Articles from the ROIS's Science Report web magazine (2017-2019)

Exploring Japan's Scientific Frontier: Connecting Data for Human Good

001–006

007–012

013–018

Focusing on Arctic,
Genome,
and Data Science

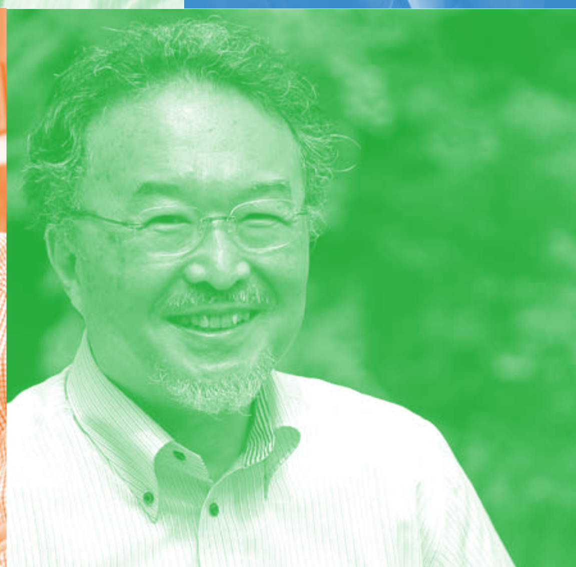
Inter-University Research Institute Corporation
Research Organization of Information and Systems









Science Report 001-018

ISBN978-4-909638-11-3 C0040

Exploring Japan's Scientific Frontier: Connecting Data for Human Good
Focusing on: The Arctic, Genomes, Data Science
SCIENCE REPORT 001-018

Author: Inter-University Research Institute Corporation Research Organization of Information and Systems (ROIS)

Editor in Chief:
The Arctic: Hiroyuki Enomoto, Professor of National Institute of Polar Research (NIPR)
Genomes: Asao Fujiyama, Specially-appointed Professor of National Institute of Genetics (NIG)
Data Science: Tomoyuki Higuchi, Director of The Institute of Statistical Mathematics (ISM)

Written by Rue Ikeya

The English version was created by Hiroko Sato, The Pitch Room, LLC (Massachusetts, USA).

Photographs: Yuji Iijima unless noted below,
Kazutaka Tateyama (p.8 left, 14-15, 25 below, 26 below),
NIPR (p.17, 18 below, 19 left, 25 above, 33 below, 37 above),
Jun Uetake/National Institute of Polar Research (p.18 above),
Jun Inoue (p.21, 22 below),
Shigeto Nishino (p29 below),
Rue Ikeya (p.33 above, 38 above, 51 below),
Hironori Yabuki (p.35),
ArCS News (p37 below, 38 below, 39 right),
Ken Kurokawa (p.43 above, p44 middle),
Primate Research Institute, Kyoto University (p.63 above)

Design: Grid Co.,Ltd., Japan

Published by Inter-University Research Institute Corporation
Research Organization of Information and Systems
Hulic Kamiyacho Building 2F, Toranomon 4-3-13, Minato-ku, Tokyo 105-0001
TEL: +81-3-6402-6200 FAX: +81-3-3431-3070
http://www.rois.ac.jp/

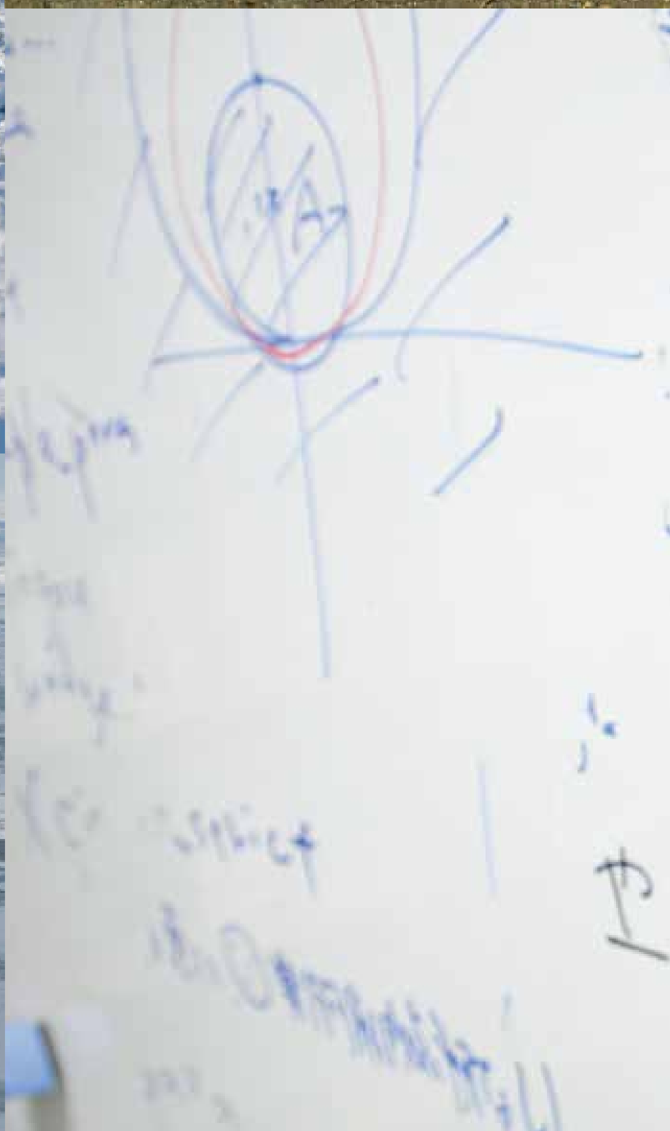
Articles from the ROIS's Science Report web magazine (2017-2019)

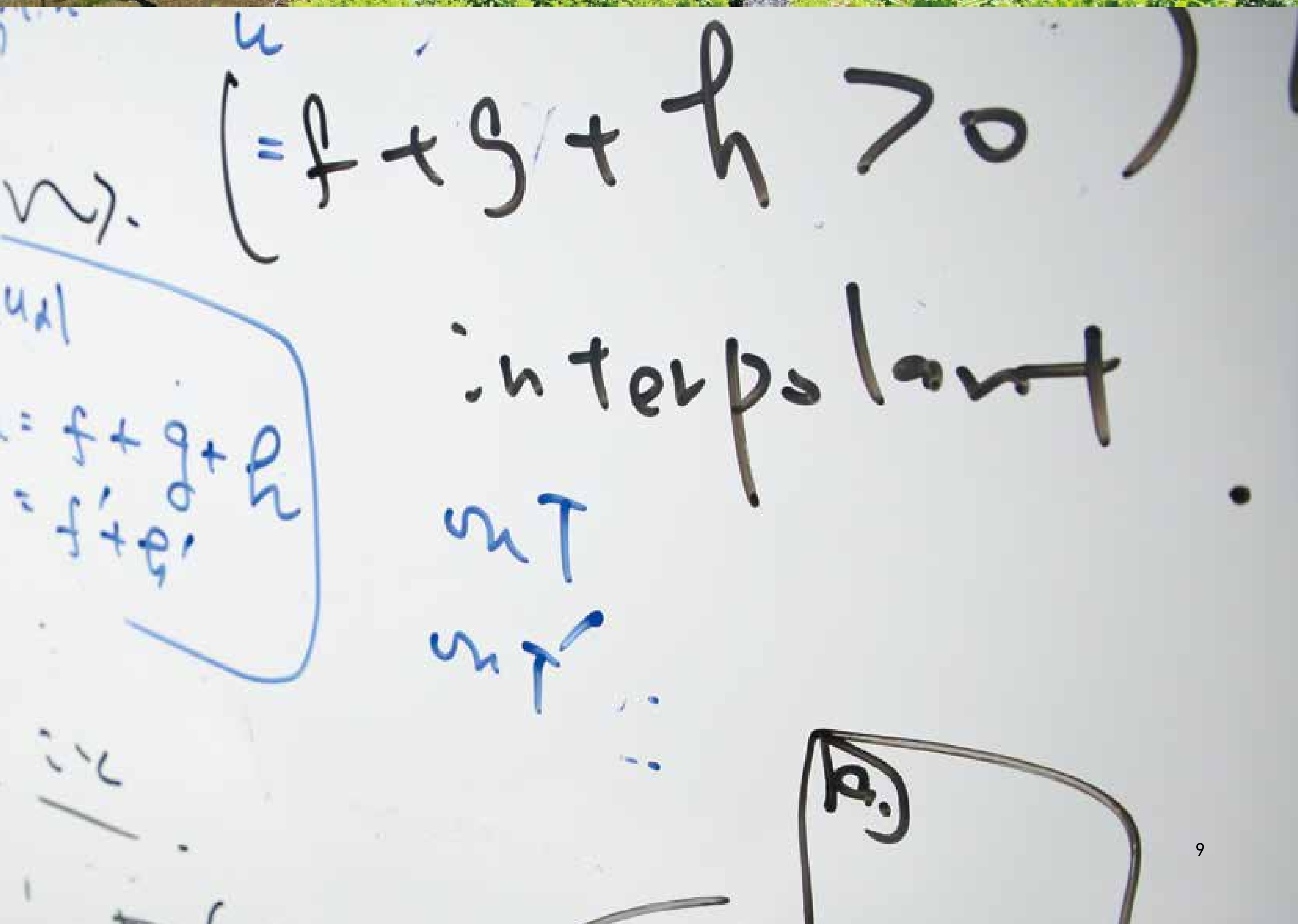
Exploring Japan's Scientific Frontier: Connecting Data for Human Good

Focusing on:
The Arctic
Genomes
Data Science

Inter-University Research Institute Corporation
Research Organization of Information and Systems

CONTENTS




interpolant

Foreword

The Research Organization of Information and Systems (ROIS) is one of the Japanese inter-university research institute corporations. It empowers the country's science community to find solutions for complex issues facing modern society.

The ROIS was founded in 2004 by four inter-university institutes—the National Institute of Polar Research (NIPR), the National Institute of Informatics (NII), The Institute of Statistical Mathematics (ISM) and the National Institute of Genetics (NIG)—as a nexus of their research network that includes all national and private universities in Japan. Utilizing the network's expansive scope of expertise, the ROIS promotes the exploration of matters related to human life, society and the environment as well as planet Earth through the lens of information and systems. While bolstering the foundation of research in each of the respective fields, the ROIS also pushes for the pioneering of new research areas and paradigms.

In addition to resource sharing and joint research within its network institutes, the ROIS promotes industry collaborations and open science for citizens' involvement. The ROIS and the four founding research institutes also strive to foster the next generations of data scientists at Sokendai (The Graduate University for Advanced Studies).

In 2016, the ROIS revamped its structure to further enhance collaborations and communications among the founding institutes and established the Joint Support-Center for Data Science Research(ROIS-DS). The Center provides support for data sharing and analysis and educational opportunities to cultivate new talents.

As the ROIS continues its effort to advance data-driven science that holds key to producing innovations in today's world, we humbly ask for your support and participation.

Ryoichi Fujii

President
Research Organization of Information and Systems

About Science Report

Science Report is a website maintained by the Research Organization of Information and Systems (ROIS) and also the name of a monthly article series that runs on the site. The ROIS, an inter-university research organization, produces the articles to keep the public informed about its latest research endeavors as well as research activities taking place at other research institutions, including universities across Japan.

The series covers diverse topics and projects, including cross-disciplinary research that employs novel research methods and concepts, and highlights how research collaborations that the ROIS works to promote help propel innovations. Offering scientific explanations in easy-to-understand language, Science Report also serves as the research community's medium through which to communicate with citizens and answer some of the most frequently asked science questions.

The ROIS periodically puts together multiple articles that were posted on the site over time in the form of a booklet. This volume, the first to be published in English, contains installments concerning three topics: Arctic science, genomics and data science.

It is our hope that this booklet will help people enjoy learning about scientists' ingenuity and visions that the ROIS works to advance.

URL https://sr.rois.ac.jp/

The Arctic as a Window into Earth's Future

Getting Ahead of Climate Change

Ask an Expert: Takashi Yamanouchi,
Professor Emeritus of National Institute of Polar Research (NIPR)

Takashi Yamanouchi, Professor Emeritus at the National Institute of Polar Research, specializes in atmospheric science and polar climatology. He serves as Editor-in-Chief of "Polar Science", an international quarterly journal. For the Summary of Outcomes, GRENE Arctic Climate Change Research Project, please go to https://www.nipr.ac.jp/grene/e/publications.html

The Arctic is becoming dirtier—or, so it may seem to people who visit Greenland to find mysterious black speckles covering ice sheet, turning the entire landscape gray.

The soot-like substance is a microbe called Cryoconite, which has spread rapidly across ice sheet cover in recent years. The emergence of Cryoconite is an example of various ecological phenomena happening in the Arctic–the place where scientists say global climate change manifests in the most extreme forms before anywhere else in the world.

"That's what we call Arctic amplification," said Takashi Yamanouchi, professor emeritus at the National Institute of Polar Research who has led Japan's major Arctic research initiative, Green Network of Excellence (GRENE) Arctic Climate Change Research Project, since its inception in 2011 through 2016.

"The effects of global warming are far more visible in the Arctic than in any other place on the planet. In fact, the average Arctic temperatures have risen twice as fast as the global average," Yamanouchi said, explaining the importance of the research initiative. And the GRENE-Arctic is now giving people a glimpse into the future, including potential uses of the Arctic Ocean in ways we've never dreamed of before.

The GRENE-Arctic is Japan's ultimate attempt at decoding climate change mechanisms. It brings together 360 scientists from 40 research institutions—including the National Institute of Polar research and the Japan Agency for Marine-Earth Science and Technology—for cross-disciplinary investigations into all aspects of the Arctic climate system from the atmosphere to biosphere.

The Arctic region refers to the area north of 66° 33' N lati-

tude, and the Antarctic region south of 66° 33’ S latitude. During the International Geophysical Year (1957-1958), there was much discussion and reporting about unusual environmental changes taking place in those regions ahead of the rest of the world. This brought polar amplification to scientists’ attention.

Recognizing the importance of the Arctic and Antarctic regions as the harbingers of global climate change, Japan soon launched observation programs in both the Arctic and the Antarctic.

In 2011, Japan established the GRENE Arctic Climate Change Research Program, calling on the country’s best talents in the science community to join in the effort to understand the process of climate change and help combat its impact on human lives.

The Cycle That Will Never End

Unlike the Antarctic that sits on land, the Arctic is mostly comprised of ocean with sea ice floating over it. This explains why the Arctic experiences the effects of global climate change faster and more severely than the south, according to Prof. Yamanouchi, who has served as the GRENE-Arctic program’s Project Manager.

“Sea ice bounces off most of the sunlight that hits its surface, but as the ice thins out, the water underneath it absorbs more light, further accelerating the warming of the water. This so-called ice-albedo feedback is the chief cause of Arctic amplification,” Yamanouchi said.

Snow and ice over the land areas of the Arctic continue to melt at a startling speed, as well, impacting the wildlife and

The scientific station of Ny-Ålesund in the Svalbard islands. The Norwegian base on the Arctic Ocean is home to 11 different countries’ Arctic observation programs. The Svalbard Treaty signed in the 1920s guarantees Norwegian sovereignty over the islands while providing its signatories with equal rights for commercial activities there. Photograph: National Institute of Polar Research

Cryoconite-covered ice sheet in Greenland. Photograph: Jun Uetake/National Institute of Polar Research

A port in Greenland. A warm current from the Gulf Stream keeps the Arctic Ocean relatively warm compared to the Antarctic. This is why some ports in such northern countries as Norway, Russia and Iceland remain open yearround without ever using an icebreaker. Photograph: National Institute of Polar Research

Arctic summer. Snow used to encapsulate the entire landscapes throughout the year. As the warming trend continues, the ground has become more visible each summer. Photograph: National Institute of Polar Research

Shuhei Takahashi of Okhotsk Sea Ice Museum of Hokkaido talks about Capt. Eiichi Taketomi's Arctic voyage during the Polar Science Symposium 2016 held on Dec. 2, 2016 at the National Institute of Polar Research

lives of the local residents.

The visible scars from the Arctic warming include the explosion of Cryoconite, the black spats found all over the surface of ice sheet in Greenland. But, Cryoconite doesn't just create an unsightly vista. The GRENE-Arctic team has observed the ice under Cryoconite melts more as it receives heat from the light-absorbing black microbe.

"We now know such changes to the ecosystem further speed up the melting of ice and snow," Yamanouchi said.

Taking Advantage of Global Warming

In the meantime, the receding Arctic sea ice may be welcome news for those in the business of international trade.

"It opens up new possibilities for ocean shipping," Yamanouchi said. "In our estimate, you can drastically shorten the travel distance between Japan and Europe by going through the Arctic Ocean instead of the Suez Canal," Yamanouchi said. "But, in order to use the route for commercial purposes, you would also have to know how much safety and financial risks there are with bypassing the Arctic Ocean. People in the industry expect scientists to figure out how to compute all that."

The GRENE Arctic successfully predicted the Arctic sea ice distribution for the summer of 2015 with little margin of error, and used the experience to draft a methodology for selecting the best ocean cargo routes. Russia, which owns oil fields in the Arctic, as well as China, Korea and other countries that have financial stakes in the region are reportedly interested in using this type of methodology to identify new and better routes.

The Heroes Who Came Before Us

Moving cargoes through the Arctic Ocean would not be an entirely new endeavor. In 1937, the Kaihomaru, Japanese agribusiness ministry's fisheries investigation ship led by Capt. Eiichi Taketomi, reached the mouth of the Kolyma River in Siberia via the East Siberian Sea, according to Shuhei Takahashi of Okhotsk Sea Ice Museum, who is knowledgeable about the topic. The ship then set out for a cross-Atlantic voyage with a year-supply of food on it in 1941 before turning around halfway through the journey. The ministry even contemplated a round-the-world trip for the ship from the Arctic to the Antarctic via the Atlantic Ocean until the world conflicts that led to the German invasion of Russia later that year interrupted the plan.

The Kaihomaru's crew for its fifth voyage included two Meteorological Observatory staff members. One of them was Takahashi's father. Shogo Takahashi, who later served as the director of the Abashiri Local Meteorological Observatory. The daily record of the voyage that Takahashi and fellow observatory staff member kept aboard the Kaihomaru became known as the "Shirokuma (white bear) Diary."

Urgent Call for Scientists

Nearly 80 years after the Kaihomaru's tours, the Arctic remains a source of wonder for scientists. Understanding what is happening at the northern end of Earth seems more critical today than ever, as urgent efforts to guard against the negative consequences of global warming continue around the world.

We believe it requires us to pull together all available resources at our disposal to tackle the threats facing the planet and its inhabitants. As Japanese researchers work together to do their part as Team Japan, we plan to share some of our key findings in this Science Report series.

Released on: May 10, 2017 (The Japanese version released on Dec. 12. 2016)

Does Arctic Warming Bring on Colder Winter for Japan?

Ask an Expert: Jun Inoue
Associate Professor of
National Institute of Polar Research (NIPR)

Dr. Inoue studies meteorology focusing on the Arctic and Antarctic atmosphere and oceans to understand how these elements come together to form polar climates. He is particularly interested in low-pressure systems that induce heat and moisture transport as well as the mid-latitude extreme weather that is driven by the decrease of Arctic sea ice. He takes advantage of observation data as well as cutting-edge techniques for data analysis for his research.

Just as global warming continues to speed up, Japan's winter is becoming colder than ever.

In 2016, Tokyo had the first snow of the season in November—the earliest snowfall recorded in 54 years. On Dec. 10 that year, 65 cm (2 ft) of snow fell in the northern city of Sapporo, followed by another 96 cm (3 ft) of snowfall in less than two weeks. The wicked weather that brought all traffic in Sapporo to a standstill remains fresh in our memory.

Anyone who gives a quick review of weather reports from around the world would realize this extraordinary wintry winter phenomenon is not unique to Japan. From North America to Europe, record-breaking temperature plunges and blizzards have been wreaking havoc across the mid-latitude area in recent years. This leads us to ask the question: Is this seasonal "cooling" interconnected with global warming that has brought on other extreme weather events such as major droughts and flooding?

While Earth Warms Up, Japan Faces Deep Freeze in Winter

In February 2015, the East River running along the island of Manhattan in New York City froze over, shutting down the commuter ferry service through the usually ice-free tidal waterway. The mercury dipped to record lows of minus 18 degrees Celsius in New York City, minus 15 degrees Celsius in the Great Lakes region and 0 degree Celsius in Florida.

"Japan had a similarly harsh winter in 2005-2006 with heavy snowfalls hammering the regions along the Sea of Japan," said Jun Inoue, associate professor at the National Institute of Polar Research. "The data shows powerful storm

systems have repeatedly developed during winter across the mid-latitude areas of the globe since the late 2000s, often resulting in heavy damage from freezing temperatures and snow."

Common sense would dictate winter gets milder everywhere as global warming escalates. Inoue said he looked at what's happening in the Arctic as a clue to understand the puzzling reality.

"I noticed that the Arctic sea ice area was the smallest on record in the summer leading up to the winter of 2005-2006," Inoue said. "I decided to look into the potential relationship between the shrinking sea ice area in the Arctic and the climate in Japan.

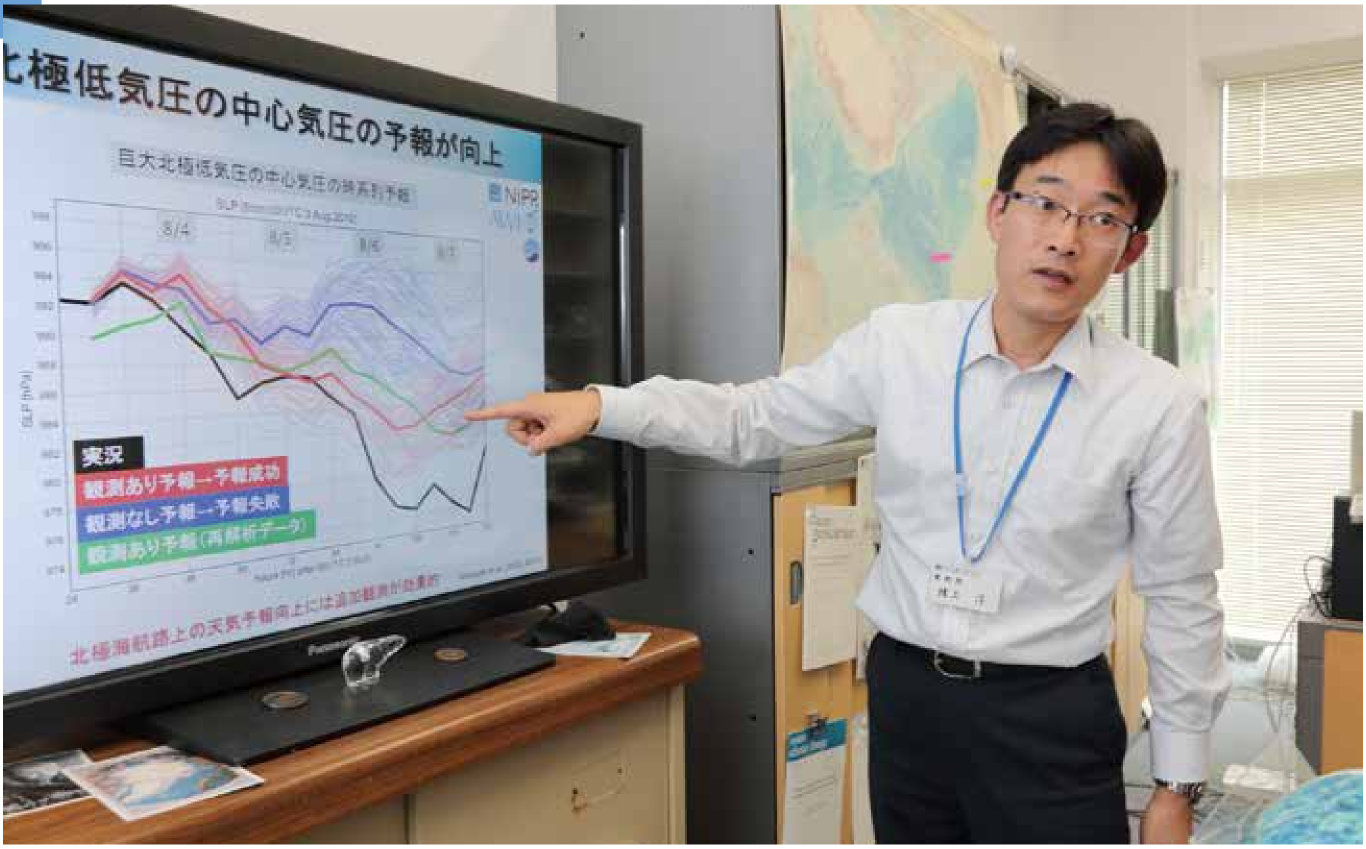
Decoding the Teleconnection Mechanism

The ongoing shrinkage of the sea ice area is one of the major environmental changes happening in the Arctic. The region has lost sea ice in the size of the Hokkaido prefecture on average every year since 1979 when data gathering via satellite began. It dwindled to the smallest size on record in 2007 before it shrunk even more in 2012. The summer of 2016 is the second on record thus far, which preceded the major cold snaps that hit North America in December that year.

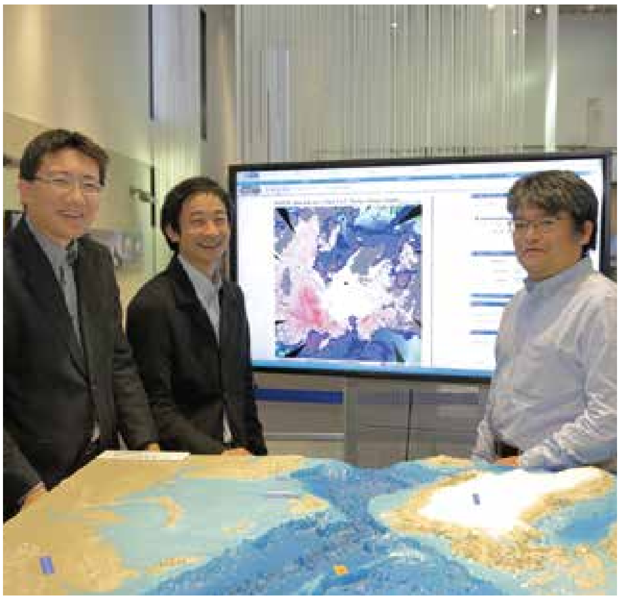
Environmentally speaking, what happens in the Arctic never stays in the Arctic. Climate around the world also changes in response to that of the Arctic, which is known as teleconnection. Now, Inoue and his team believe they have decoded the mechanism of teleconnection.

"The Gulf Stream is becoming warmer and shifting northward, and this is keeping the high-latitude areas in Europe

The Sapporo city was already deep into the snow season as of Dec. 20, 2016. Photograph: Jun Inoue



A radiosonde released into the air for aerological observation. Helium-filled balloons are used to lift radiosondes from more than 600 locations around the world twice a day. The small box-shaped device that functions as a weather station contains instruments for measuring atmospheric pressure, temperature, humidity, wind direction and speed and other parameters and transmits data by radio to a receiving station. "It is an old-fashioned observation tool. But, when used simultaneously at many locations, it helps you understand what is happening in the atmosphere," Inoue said. Photograph: Jun Inoue



Research fellows Takeshi Terui (from left) and Takeshi Sugimura and Research Associate Professor Hironori Yabuki, all of the National Institute of Polar Research, are developing the Arctic Area Data Archive System (ADS) shown on the computer monitor in the photo. The ADS processes data in real time to map out the Arctic sea ice area, simulates the best route through the Arctic Ocean, provides data on Arctic sea ice concentration and sea surface temperature and other information useful to Japan and the surrounding countries. The system launched in 2012 as a way to bring together the data on all aspects of the Arctic climate and environment as well as the space data compiled by the Japan Aerospace Exploration Agency (JAXA) and offer it to the public in easy-to-understand visual formats. Boasting one million page views per month, the ADS now serves as the platform for the ArCS team to share its research results with such international organizations as the Global Earth Observation System of Systems (GEOSS) and the WMO.

warmer in winter. We found that this also allows the low-pressure system that develops above the North Atlantic to move farther north toward the North Pole. Southern winds from this low-pressure system also push sea ice farther north," Inoue explained.

Heat pours into the Arctic, accelerating the warming of the Arctic Ocean. In the meantime, the northward shift of the low-pressure systems results in the development of high-pressure systems above the continent and this leaves there cooler.

"Japan lies in between high-pressure systems to the west and low-pressure systems to the east during winter. These western high-pressure systems above the Asian continent that give us Old Man Winter" Inoue said.

Inoue continues to investigate this mechanism through an Arctic region research project, the Arctic Challenge for Sustainability (ArCS).

Understanding of Global Climate Requires Global and Reliable Data

Comprehensive climate studies require different types of data from various points around the globe. Inoue said collecting atmospheric data, such as air temperatures and humidity levels at up to 30 km above the ground is as critical to weather and climate studies as ground- and satellite-based observation data. Observation stations that can gather such data are mainly located in the northern hemisphere, however. There need to be more observation stations in the Arctic and Antarctic regions as well as other underpopulated parts of the world, Inoue said.

Collected data benefits not only climate researchers but also ordinary people as weather prediction centers in Japan, the U.S., Europe and many other countries use it every day. The World Meteorological Organization (WMO) has a data exchange network that connects observation centers and weather prediction centers around the world. Observation stations first send raw data to the WMO, which log and distributes it to the network members. The centralized system ensures weather prediction centers and climate researchers have real-time access to trustworthy data from a reliable source, Inoue said.

More Arctic Data, the Better Forecasting

The speed and accuracy of weather forecasting today has a lot to do with the use of the real-time observation data coming out of the WMO.

"Real-time data enables you to compare the accuracy of various forecasting models. You can also plug in different data to see how that affects scenarios. Using real-time data this way helps improve the accuracy of weather and marine forecasting greatly. It also helps you understand what other data should be gathered to make forecasting even better," Inoue said.

"Climate researches in Japan have traditionally focused on the tropical climate, but more researchers are beginning to focus on the Arctic region these days," Inoue said. "Both 2017 and 2018 are slated as the Year of Polar Prediction. I am sure such a heightened interest in the Arctic climate will help improve the quantity and quality of data we gather and push the advancement of weather forecasting techniques."

Released on: Jun 27, 2017 (The Japanese version released on Jan. 10. 2017)

What is Happening with the Arctic Sea Ice?

Ask an Expert: Kazutaka Tateyama
Associate Professor of
Kitami Institute of Technology

Dr. Tateyama is one of the few scientists in the world conducting research on the changes of sea ice volume in the Arctic and Antarctic oceans and the Sea of Okhotsk. He is working to develop a system that pulls satellite observation data to estimate the thickness of sea ice, which is currently impossible to do without getting hands on the ice. While monitoring climate changes related to sea ice, Dr. Tateyama helps lead research projects on Arctic and Antarctic developments, including sea route and coastal oilfield developments in the Arctic Ocean. He participates in the GRENE Arctic Climate Change Research Project and the Arctic Challenge for Sustainability (ArCS).

If you think seawater should taste pretty much the same everywhere around the world, think again. In the Arctic, the salt content of the surface seawater has always been below the global range of 3.1 to 3.8 percent, and it's becoming even less salty in recent years.

Experts point to the increasing amount of water flowing into the Arctic Ocean from the surrounding rivers as the culprit of the reduced salinity. The diluted seawater doesn't freeze the way saltier seawater does. That sparks a chain of events that can exacerbate global warming.

From salinity to chemical balance, all data related to the Arctic water cycle provides clues for understanding the intricate workings of the Arctic and global climate systems.

Now, we would like to invite you to take a closer look at what's happening with the Arctic sea ice and how that helps fuel the climate change with the help of our expert.

Distinguishing Between Sea Ice and Iceberg

In the Arctic, ice comes in all shapes and forms. Sea ice refers to frozen seawater with thickness ranging from a few tens of centimeters to 10 meters. An iceberg, on the other hand, is compressed snow layers that forms on a glacier or an ice shelf and breaks away from it after repeatedly washed by tides. It floats into the ocean like the one that crashed into and destroyed the Titanic. Icebergs are particularly abundant in Central Greenland where ice mountains of various sizes continuously break away from the shorelines that look up to 2,000- to 3,000-meter-high ice sheets.

"You listen to the sounds of the vessel hitting and crushing the ice below while moving through the Arctic Ocean, and

you know how old the ice is," says Dr. Kazutaka Tateyama, associate professor at Kitami Institute of Technology. "New ice under a year old sounds like shaved ice. Ice that's gone through two or more summers makes thumping noises or squeaks like pieces of Styrofoam rubbing against each other."

Sea ice and icebergs look alike from the above because they are both covered in snow.

"You need to look at the cross-section of ice. Newly formed ice looks greyish and you can see seawater through it. One- to two-year-old ice has blueish tones to it, and it turns greener as it ages."

Dr. Tateyama studies Arctic sea ice with a focus on ice volume changes. He says local animals often approach the research crew during the fieldwork. They include polar bear, which use sea ice as a foothold to hunt seals. For polar bear, sea ice plays a crucial role in their survival, Dr. Tateyama says.

"Some years, I run into polar bear every day. Some other years, little sea ice would be left, and bear don't come, either," he says.

Heavy Brine Makes Seawater to Circulate Around the World

The lower the salt content of water is, the faster it freezes. When new sea ice begins to form, a matrix of freshwater ice crystals develops. These ice crystals contain pockets of high salinity sea water known as brine cells in them. The freezing temperature in the Arctic is approximately minus 2 degrees Celsius (28.4 degrees Fahrenheit). This means brine hardly ever freezes and only grows more concentrated and heavier, sinking towards the bottom of the ocean to become part of the "global ocean conveyor belt"—deep seawater that circulates around the globe. Temperature and salinity along with energy from the wind received by the surface seawater drive this down-reaching current southward across the Atlantic Ocean and then eastward to the Pacific Ocean past the north of the Antarctic. Deep seawater then moves up and down the Pacific and returns to the Atlantic Ocean. The round-the-world journey of deep seawater takes about 1,000 years to

Polar bear that showed up while the research crew was gathering data.
Photograph: Kazutaka Tateyama

Various types of sea ice seen near the shore of the Sea of Okhotsk. Ice that's been just formed (upper left) is thin and semi-transparent and gradually grows thicker (upper right and lower left). When pieces of new ice collide with each other, their soft edges curl up and begin to look like lotus leaves (lower right). But, the drifting patterns of sea ice in the Sea of Okhotsk have largely changed over the past 25 years, and the volume of sea ice is also declining. Sea ice drives minerals and all the things that enrich the sea water to circulate across the oceans, creating great fishing grounds. Sea ice, which covers the sea surface like a lid, also helps calm the oceans. However, sea ice hasn't been able to perform such functions as effectively in recent years, according to Dr. Tateyama. "Calm oceans may turn rough in the future," Dr. Tateyama says. Photograph: Kazutaka Tateyama complete.

The deep ocean circulation shows how heat transfer and salinity balance of snow and ice in the Arctic contribute to climate events a halfway around the world.

"Over the past two to three decades, which is a short term when you are discussing climate, we have been seeing more and more thin ice and less aged ice in the Arctic. This is a direct result of global warming," Dr. Tateyama says. "Thin ice quickly melts in summer heat, leaving a larger section of ocean surface directly exposed to the atmosphere. This causes more seawater to evaporate, triggering heavier and more frequent rain and snow. The salinity of the Arctic surface seawater is declining as the volume of water streaming into the ocean is increasing. This may cause the Arctic to regain some of the lost sea ice area because lower salt content helps water to freeze. Sea ice extent may also grow if a lot of snow falls on the ocean and freezes.

Such seemingly "cooling" phenomena may not last too long, however, as they are symptoms of global warming. The Syowa Station in Antarctica observed the fast ice at record thickness recently, and this can also be explained by global warming.

Decoding the Arctic Climate Through On-site Verification of Aerial Monitoring Data

Dr. Tateyama and his research team visit the Arctic every September or October to find out what's happening at ground zero of climate change. Sea ice is a relatively new area of study led by Japan's Tateyama team. Experts believe obtaining accurate and detailed information about sea ice can reveal how climate and those which live in it are intertwined. Such data can also help improve weather forecasting, along with other benefits. This is why the global science community is paying close attention to Dr. Tateyama group's research projects.

For the monitoring of sea ice, researchers use remote-sensing instruments that do not damage the environment, such as electro-magnetic inductive device and microwave radiometer. Dr. Tateyama is experienced with the technology as he has used it since 2003 when he was a university student.

"We want to be able to not only remotely detect ice and its age but also its thickness preferably in 10 cm increments. We are trying to develop a remote-sensing system for that," Dr. Tateyama says. "We also hope our use of numerical models for the analysis of satellite and field data will help us overcome some challenges, such as dealing with multi-layered, oddly shaped ice and distinguishing between seawater and a pool of water on sea ice."

Dr. Tateyama is unique in that he combines surface and aerial observation data to improve the data accuracy for to obtain more accurate and precise data.

"In April 2016, our surface observation team drove dog sleds along Greenland's shores to stake flags in the fast ice there while our aerial team followed the markings by helicopter to gather data.

The Japanese research group conducts various research projects in the Arctic. Prof. Atsuko Sugimoto of Hokkaido University has spent the past 20 years monitoring sub-Arctic forests with a focus on the cycles of carbon and water and their impact on global warming. "Larch is highly adaptive in the high-latitude areas of Northeast Eurasia. There are even forests solely comprised of one type of larch. Human activities impacted the climate over the past 100 years or so. When you look at the tree rings, you can tell what has been happening to the environment over those years," Sugimoto says. Permafrost open land dotted with taiga spreads to the north of the forestland. Tundra where only moss grows can be found farther north near the Arctic Ocean. "We are researching the ecosystem in the Russian city of Chokurudakh (70° N) that saddles over taiga and permafrost areas. We are keeping an eye on the taiga forests to see if it expands northward due to global warming," Sugimoto says. Sugimoto is also monitoring how thawing permafrost may be affecting the water cycle and how much methane is being released from the wetland. "Thawed permafrost turns into wetland, which releases methane. But, the forest absorbs methane. We are continuing with the observation at our station in Chokurudakh (70° N) and Yakutsk (62° N), as well, so we can find out what will happen to an area that has both forests and wetland," Sugimoto says.

Dr. Tateyama's research project is a collaborative one with the Arctic Challenge for Sustainability (ArCS), and his team closely works with the ArCS team led by Arctic glacier experts.

"We are focusing on the points of contacts between the land and sea. We research the water cycle and mass balance of glaciers that flow into the ocean and dissolve into the seawater. Much of what we do has never been tried before, and so it's very exciting. We are hopeful that a lot of useful information will come out of the work," Dr. Tateyama says.

Released on: July 10, 2017 (The Japanese version released on Jan. 20. 2017)

Will Living Organisms Survive the Arctic Climate Change?

Ask the Experts: Dr. Shigeto Nishino and Dr. Eiji Watanabe, Japan Agency for Marine-Earth Science and Technology (JAMSTEC)

Dr. Nishino (left) leads Arctic observational projects aboard the JAMSTEC's research vessel, the Mirai, for collections of oceanographic and atmospheric data. He specializes in the study of how global warming and declining sea ice affect the physical and chemical conditions of the Arctic Ocean and its ecosystem.To the right of Dr. Nishino in the photo is Dr. Eiji Watanabe, a researcher at the Institute of Arctic Climate and Environment Research (IACE), JAMSTEC.

In the kingdom of marine life, the food chain starts with barely visible organisms known as plankton. Small fish devour them, and larger fish feast on small fish. If plankton vanish, all the rest higher up in the pecking order would starve to death.

Some parts of the Arctic now face the danger of losing plankton as sea ice continues to melt, diluting nutrients in the seawater to make it inhospitable for plankton. The ice melt is also accelerating ocean acidification, resulting in a harsher living environment for crustaceans like crabs and shellfishes.

The world's oceans are a boundless, 3,800-meter-deep expanse with an intricate ecosystem in which its inhabitants and their surrounding environments, from temperature to salinity levels to concentrations and ratios of chemical components, are intertwined with each other. And, in order to understand how these oceans may change due to global warming, we need to first look at the Arctic Ocean as it is one of the areas where most dramatic environmental changes are occurring in the worlds and its conditions have cascading effects on the entire planet.

So, what is really happening inside the Arctic Ocean? And, what will it mean for the world's oceans that cover 70 percent of the Earth's surface?

Experts from the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) will take you on an Arctic journey for answers.

Heading for the Arctic Aboard *Mirai* the 8,706-ton Research Vessel

The *Mirai*, an 8,706-ton research vessel owned by the JAMSTEC, is one of the world's largest of its kind. It also serves as Dr. Nishino's research base and home for months on end every year.

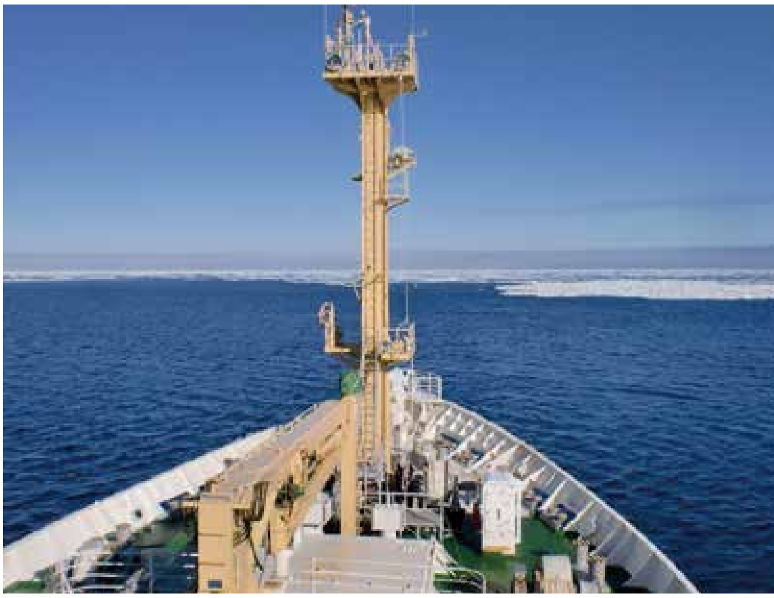
The annual voyage that starts and ends in Japan always takes Dr. Nishino and his colleagues through the North Pacific toward the northern Bering Sea first.

"The northern Bering Sea is a treasure trove of marine species, and so we focus on documenting about them as much as possible while moving north," Dr. Nishino says.

Once through the Bering Strait that separates Alaska and Siberia, the vessel would enter the Chukchi Sea, a biodiversity-rich area in the Arctic Ocean. The *Mirai* would then head into the deep Canada Basin.

"This is where some major environmental changes have been happening in recent years," Dr. Nishino says of the basin. "Sea ice has been melting, and the ocean surface temperatures are rising. We see the seawater becoming more diluted with freshwater, which lowers the nutrient concentration of the water and accelerates the ocean acidification. We gather a wide range of data in this area to keep track of these changes from a macro viewpoint," says Dr. Nishino, who has also served as the chief scientist for the Arctic research voyage organized as part of the Arctic Challenge for Sustainability (ArCS) in 2016, as well as the previous Arctic cruises in 2013 and 2015.

In addition to oceanographic and atmospheric data, the scientists onboard also collect information related to waves



The edges of sea ice near Point Barrow in Alaska (71°25'N, 158°40'W) as seen from aboard the Mirai. Photograph: Shigeto Nishino

During the JAMSTEC's 2016 Arctic voyage, a science team successfully used a prototype of a small autonomous underwater vehicle (AUV) to gather data, such as water temperature and salinity, and to capture images of the underside of sea ice as well as the activities of plankton below sea ice. Since the late 1990s, the JAMSTEC has strategically installed measurement devices undersea. These instruments include current meters and water temperature and salinity sensors, all of which are retrieved after a period of time. The agency has also recently begun mooring biochemical sensors for the measurement of oxygen and phytoplankton in seawater.

and droplets, greenhouse gas, marine physics, the ocean's chemical environment, and marine life and benthic organisms ranging from plankton to seabirds.

"Though the *Mirai* is capable of maneuvering near ice-edge, there are times when sea ice blocks us from what we want to see or measure, or when weather conditions make data-gathering impossible. So, we hold a meeting every day to create a 'tomorrow menu' for the next day," Dr. Nishino says. "Our goal is to use the *Mirai* to capture the changes occurring in the Arctic Ocean to the best of our ability."

More freshwater for More Acidic Seawater with Less Nutrients

The Arctic Ocean has two major sources of water: The Pacific Ocean, from which seawater flows in through the Bering Strait, and the Atlantic Ocean. The Pacific water, which is rich in nutrients and plankton, helps enrich the Arctic marine ecosystem.

In recent years, sea ice is melting dramatically in the Arctic Ocean. This means a large amount of freshwater in the Arctic is turning from solid to liquid, diluting the seawater in the surface layer. The nutrient concentration in the surface layer declines (oligotrophication), and the Arctic Ocean's alkalescent seawater becomes more acidic. These changes all could negatively impact the Arctic marine ecosystem.

But, the domino effect of melting ice doesn't stop there. Once sea ice is gone, it directly exposes the seawater to the atmosphere. That accelerates the absorption of carbon dioxide in the air into the ocean, which, in turn, exacerbates the acidification of the seawater.

"The Arctic Ocean acidification is advancing at a much faster rate than other oceans are," Dr. Nishino says.

Changing Oceans as Carbon Dioxide Eater

Oceans absorbs carbon dioxide like a sponge. In fact, one-thirds of anthropogenic carbon dioxide emitted into the atmosphere disappears into the oceans.

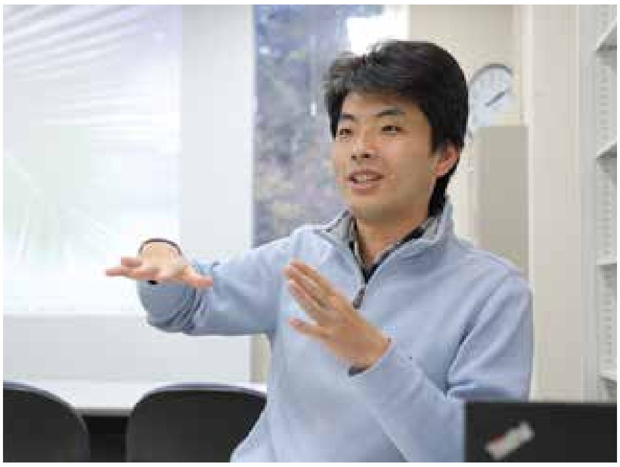
When carbon dioxide enters the seawater, phytoplankton that floats near the ocean surface takes in some of them in the process of photosynthesis. Zooplankton eat phytoplankton, which feces and bodies sink to the seafloor to become sediments. This carbon sequestration system known as the "biological pump" effectively removes carbon dioxide from the ocean's surface and stores it at the bottom.

The problem is, Dr. Nishino says, phytoplankton that plays a crucial role in the biological pump, cannot thrive without nutrients in seawater.

"As the Arctic seawater gets more freshwater mixed in, the nutrients in the surface water become diluted. Plankton populations in the Arctic decline as a result, making it harder for the biological pump to function effectively," Dr. Nishino says.

Acidification of the seawater might also make the Arctic uninhabitable for organisms that have calcium carbonate shells, including plankton, crabs and shellfishes, as their shells could melt away. In science, calcium carbonate saturation state is expressed with the symbol of Ω. Anything lower than $\Omega=1$ is "undersaturated," meaning calcium carbonate in shells and bones of living organisms can seep out into the seawater.

"We began to see Ω values for the Arctic surface water dip below 1 in the early 2000s. But, it doesn't mean shells of plankton, shellfish, crabs and things like that will melt right away. It requires more research to understand how the marine life may adapt to their changing environments," Dr.Nishino says.



Dr. Eiji Watanabe is responsible for creating high-resolution oceanography simulations based on the data collected by Dr. Nishino and his colleagues. He uses the JAMSTEC's supercomputer named Earth Simulator and runs two existing models on it: COCO, which allows simulation of ice melting and current movements, and NEMURO for simulation of the marine food chain that includes nutrients, phytoplankton and zooplankton. COCO was developed by the Center for Climate System Research, The University of Tokyo (currently the Atmosphere and Ocean Research Institute, The University of Tokyo), and NEMURO by the members of The North Pacific Marine Science Organization (PICES). "We now have better technologies to allow for interdisciplinary simulations that takes physical, chemical and biological processes into consideration. Also thanks to the improvements in computer performance and simulation models, we can visualize anything from a 10-kilometer-radius eddy in the ocean to the entire Arctic ocean circulation with a horizontal resolution of 5 kilometers. This was unimaginable just 10 years ago. We can use these simulation models to analyze how nutrients and plankton move through the ocean and conduct research on algae that grows under sea ice," Dr. Watanabe says. JAMSTEC researchers are also working with their global counterparts to improve 4D simulation models of the Arctic marine food chain and have so far made great strides in the project.

Released on: Aug. 10, 2017 (The Japanese version released on Feb. 10, 2017)

What Is the Arctic? It's Complicated.

Ask an Expert: Shinichiro Tabata
Professor of Hokkaido University

Prof. Tabata is a renowned economist with expertise in statistical and comparative analysis of Russia's economic growth. His research topics include the Russian gas and oil industry's impact on the country's gross domestic product (GDP). He has served as the director of the Slavic-Eurasian Research Center at Hokkaido University through January 2017, and has supported the center's research projects on inter-regional comparative studies, border research and indigenous studies as well as the ArCS' interdisciplinary research on the Arctic region. He currently serves as a professor at the Slavic-Eurasian Research Center and the university's Arctic Research Center.

How do you define the Arctic?

The answer to the question depends on whom you ask.

Russia might describe it as the country's new frontier for oil and gas. Norwegians would see it as a destination for tourists wanting to marvel at the aurora-lit skies. In the case of the indigenous people like the Inuit and the Yupik, the Arctic is also ground-zero for their survival against the effects of climate change.

The truth is, the Arctic is all the above and more. The region that falls inside the Arctic Circle–an imaginary circle around the North Pole located at about 66° 33' North–comprises the high seas as well as land and seas belonging to eight different countries, ranging from the U.S. to Canada to Russia to Scandinavian nations. The culturally diverse group has worked together on common causes, but their competing economic interests also have brought some of them to a negotiation table from time to time.

The political landscapes surrounding the region will likely grow even more complex as an increasing number of non-Arctic nations are eager to become an Arctic partner in the developments of a new sea route and natural resources as well as for the protection of the environment.

So, what does the Arctic really mean to these stakeholders? And, how do those living in the Arctic Circle feel about it?

Prof. Shinichiro Tabata of Hokkaido University will share his outlook for the Arctic's future from the social sciences standpoint.

One of the positive consequences of the end of the Cold War may be the collaborative relationships forged amongst the Arctic nations that share common goals. In 1996, the Arc-

tic Council was established with eight countries that have jurisdiction over land and/or seas located within the Arctic Circle as its members. The members are Canada, the Kingdom of Denmark, Finland, Iceland, Norway, the Russian Federation, Sweden and the United States of America. The Council acts as the intergovernmental forum to promote "cooperation, collaboration and interaction among the Arctic states, Arctic indigenous communities and other Arctic inhabitants on common Arctic issues," particularly sustainable development and environmental protection of the Arctic, according to the Council. Six organizations representing the indigenous peoples of the Arctic also hold "permanent participants" status.

"Among all the members, Russia by far has the most say," Prof. Tabata says. "Russia accounts for 60 percent of the economic activities in the region. The existing Arctic sea route also lies to the north of Russia. I bet Russians think the Arctic Ocean belongs to their country."

Humanities and Social Sciences Through Arctic Lens

The Slavic-Eurasian Research Center at Hokkaido University, where Prof. Tabata works, has conducted research on Russian and Eastern European people and societies for more than 60 years. In 2014, the center added "Eurasia" to its original name of The Slavic Research Center to reflect the range of geography and ethnicity encompassed in its research.

"Most people think 'Slav' means Russian descent. In reality, though, there are many nationalities or ethnic groups in former Soviet Union and Eastern Europe," Prof. Tabata says. "Hungary, Romania, Estonia, and Central Asian countries and Caucasus are typical examples. Our center's studies now include Mongolia and Greenland, as well."

Hokkaido University is known for its cultural and anthropology research, which includes studies of Siberia and indigenous people, such as the Ainu. Prof. Tabata says humanities and social sciences studies through the Arctic lens can offer interesting insights.

"The Sami people in Scandinavia and the Ainu have things in common culturally, which I think stems from them having had to adjust to the cold climate," Prof. Tabata says. "But, the traditional area study framework doesn't allow us to bunch them together as the people of the Arctic. So, I take a differ-

Tromso, an island located at 69 degrees north latitude and 350 kilometers (217 miles) north of the Arctic Circle, is a popular destination for tourists looking to experience Arctic natural phenomena, such as midnight sun and the aurora.

The room dedicated to the Arctic Research Center's exhibit at The Hokkaido University Museum is open to the public for free and draws many visitors year around. (For the museum's website, please go to https://www.museum.hokudai.ac.jp/english/.)

ent approach to research, which is to look at all areas within the Arctic Circle, from Canada to the U.S. to Russia to the Scandinavian countries, as one region. Speaking of which, I believe sustainability is the biggest issue facing the Arctic region as a whole from the economic and social standpoint," Prof. Tabata says.

Russia's New Frontier for Oil and Gas

The world is growing keenly aware of how much natural resources the Arctic region has to offer. Fish are just as abundant off the coast of Northern Norway and in the Icelandic coast as in the Bering Sea. Northern Alaska accommodates crude oil fields. Russia and Norway have often negotiated their rights over the enormous undersea oil and gas reserves spanning both countries' territories.

The region's oil and gas reserves are particularly drawing the world's attention as Russia tries to further tap into them.

"Russia ranks the number one or two in the world in terms of oil and gas production and exportation as well as their reserve amounts. And so, the petroleum industry is extremely important to the country's economy," Prof. Tabata says. "In fact, it was the petroleum resources in Baku on the Caspian Sea that propelled Imperial Russia's industrialization in the late 19th Century. Then, Volga-Ural emerged as the "second Baku" in the 1940s and became the hub of the Russian oil and gas production before Western Siberia took over the spot. Production is plateauing at both locations now, which is why the industry is migrating farther north and east for exploration and drilling. Given the industry's significance to the country's economy, the Arctic development isn't just important; it's a necessity for Russia."

Putting Yamal on the World Economic Map

Among all the areas being developed for potential production of oil and gas, Russia is particularly invested in the Yamal Peninsula. Yamal means "the end of the world" in the local language. Permafrost covers most of the peninsula, where indigenous people maintain their traditional nomadic lifestyle by herding reindeer.

"Since 2007, the government has been working on the project to construct a liquefied natural gas production facility here to take advantage of the large oil and gas reserve. They are constructing a seaport, as well. Once a new Arctic sea route opens, the port will allow exports to Europe year around and to Asian countries, such as Japan and China, during summer," Prof. Tabata says.

But, to get to the point, Russia need to rely on other countries for cutting-edge deep-sea drilling and LNG production technologies as the country lacks such expertise required for the development.

"American capital has been involved in Russian oil and gas exploration since the Baku days. Russia couldn't develop any additional oil or gas fields without the technological support from the U.S., Japan, France and the Netherlands, among other countries. Russia also needs Asian countries that import its petroleum products to show interest in the project," Prof. Tabata says.

Hironori Yabuki, project associate professor at the National Institute of Polar Research, studies frost of Northeastern Eurasia. Frost refers to ground soil that stays at or below the freezing temperature. Permafrost refers to ground soil that remains in the frost state for 2 or more years. Permafrost spreads out from Siberia and covers 24 percent of the Earth's surface. Global warming causes thawing of frost, which in turn helps accelerate global warming. In the upper left and right photos, Prof. Yabuki and his team install an earth thermometer after boring a 60-meter hole in the ground in Tiksi in a Tundra on the Arctic shore in 2006. Data collected is expected to enhance understanding of different stages that permafrost and terrestrial vegetations go through as they change amid global warming. "Communications with local people is paramount for success of onsite investigations and monitoring," Prof. Yabuki says. The lower left photo shows methane gas coming up from a pond that was once frost in Yakutsk in Eastern Siberia . The lower right photo shows frost thawing in a town near Yakutsk in Sakha Republic of Northeastern Russia, which exposes utility poles (seen in the background) to the danger of falling. (Photographs: Hironori Yabuki)

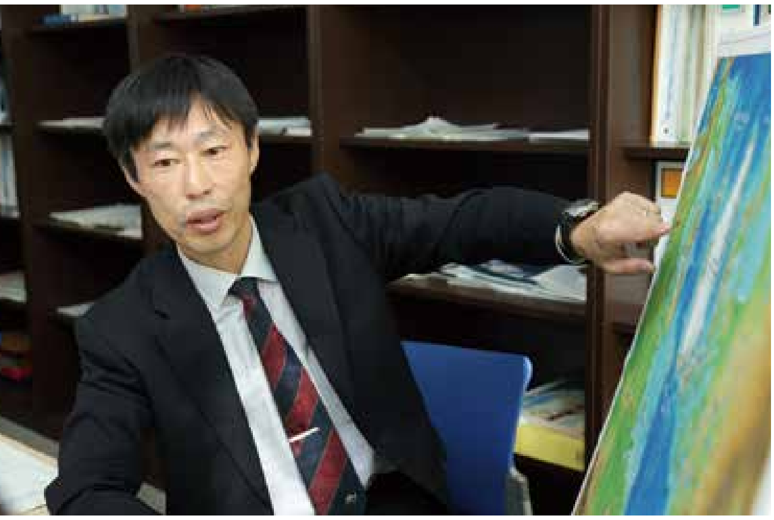
Japan has served on the Arctic Council as an observer since 2013 when the Council granted the status at its eighth ministerial meeting. As of July 31, 2017, there are 13 non-Arctic states holding observer status, including such Asian countries as the People's Republic of China and the Republic of Korea.

"The Arctic has precious metals and other valuable natural resources that are important to the Japanese economy. Another and more crucial reason for Japan being involved in the region is that Japan is on the future Arctic sea route," Prof. Tabata says. "Also, given China's and Korea's strong interest in the Arctic, it is imperative for Japan to maintain a presence on the Council."

If all these countries' attention is any indication, the Arctic is on its way to becoming a major stage for international politics, business and science.

Released on: Aug. 21, 2017 (The Japanese version released on March 10, 2017)

Team Japan Looks to Connect All the Climate Dots in the Arctic



Ask an Expert: Hiroyuki Enomoto
Professor of National Institute of Polar Research (NIPR)

Dr. Enomoto serves as professor at the NIPR and the director of the NIPR's Arctic Environment Research Center (ÆRC). He holds an undergraduate degree from Hokkaido University and received his PhD in physical geography from Swiss Federal Institute of Technology (ETH) in Zurich for his study on sea ice and climate. Prof. Enomoto has participated in various sea ice, glacier and cryosphere observatory programs in the Arctic and Antarctic and joined the NIPR in 2011 following his tenure at Kitami Institute of Technology in Kitami, Hokkaido. As sub-project director of the ArCS, which was launched in 2015, Prof. Enomoto oversees the maintenance of global research and observation sites in the Arctic and assembles a delegate of experts to the Arctic Council related meetings to promote Japan's "sci-tech diplomacy" efforts.

For many Japanese scientists, the "end of the earth" is becoming a familiar corner of the earth.

Recognizing the urgent need to tackle climate change, Japan named its first-ever Arctic ambassador in 2013, followed by the Arctic Council's admittance of Japan as one of its observer states. By late 2015, the country would develop "Japan's Arctic Policy," spelling out its commitment to Arctic monitoring and research as well as collaborations with the international research community. With hundreds of Japanese scientists working together under the framework of the Arctic Challenge for Sustainability (ArCS)–Japan's national flagship project for Arctic research–the Arctic no longer seems like a faraway place on the end of the earth. It's simply their research base.

Now, global leaders are hoping these Japanese scientists will help the world make sense of climatic changes in the Arctic with their data and analysis and address society's needs. So, how is Team Japan trying to do just that? An expert from the National Institute of Polar Research answers to the question.

Japan's Role as Expert Witness for Arctic Climate Change

Scientists are a hot commodity in global political scenes these days. Every stakeholder participating in such events as the Arctic Council's meetings wants scientists by their side while making their points.

"Arctic climate change has global consequences, and people want to know why they are experiencing extreme climat-

ic phenomena and what it means for the future. If you have ample, concrete scientific data to back up your policy proposal, more people will listen to you," Prof. Enomoto says.

That's where Japanese scientists come into the picture.

"People want scientists to find ways to use their research results in meaningful manners for the benefit of the society. I believe that's the mission of the 21st Century Arctic research," Prof. Enomoto says.

Prof. Enomoto points out, however, that the Arctic climate is changing so fast that the problems that the scientists are trying to tackle can be moving targets.

"Nature is moving fast, and it's like humans are trying to catch up with it. But, nature continues to beat us to it," Prof. Enomoto says.

This may mean that we, as an international research community, need to decide what is the most important thing to monitor, and develop "observation strategies."

The Arctic Challenge for Sustainability (ArCS) aims to decode Arctic climate change by bringing together all the experts and resources that Japan has to offer on the topic just as its predecessor, the Green Network of Excellence (GRENE) Arctic project, did (See Science Report 001). This "Team Japan" approach is drawing the attention of the global research community, according to Prof. Enomoto, who serves as part of the ArCS core leadership.

"People look at us as this compact team from a mid-latitude country going all out to advance the Arctic research," Prof. Enomoto says. "Just because you are an Arctic country, that doesn't necessarily mean you have the best know-hows for observing the region's climate. Strategic selections of

Japanese scientists are becoming more visible on the global stage. Some of them, photographed here, participated in the September 2016 general assembly of the Protection of Arctic Marine Environment (PAME), which was held in Portland, Maine, in the U.S. The PAME is an Arctic Council working group for the conservation of the marine environment and non-emergency pollution prevention. Photograph: Natsuhiko Otsuka (From ArCS News, ArCS website)

Prof. Enomoto talks to the audience in the ArCS public seminar, "The Arctic Research and Japan–What Motivates Us to Study the Arctic," held on March 18, 2017, at Seiryo-Kaikan in Chiyoda-ku, Tokyo. The sea ice area in the Arctic shrinks to a minimum size every September. The minimum size from 2012 is the smallest on record thus far, followed by that of 2016. The sea ice area has not grown back at a normal speed since September 2016, however, and the recovery has been particularly slow off the coast of Norway and near Alaska, according to Prof. Enomoto's latest findings. "Arctic research is like a mountain. You think you've climbed quite a bit and then find out you were still at its base. This is happening more as the Arctic climate continues to change," Prof. Enomoto says.

Qaanaaq, a coastal village in northeastern Greenland, is among the communities that host the ArCS's observation base. (Right) Photograph: Yoshihiko Ohashi (From ArCS News)
Field observation on the Bowdoin Glacier in July 2016. The site is 20 kilometers from Qaanaaq, and the glacier is changing rapidly especially on its outer edges. Team Japan has gathered data here since 2013. Photo: Evgeny Podolskiy (From ArCS News)

The ArCS's public seminar for the residents of Qaanaaq in July 2016. The seminar was held with the help of an interpreter who speaks English and Greenland's local language. In the Greenland which has a population of 56,000, the local language is the official language. (Photograph: Bungo Nishizawa From ArCS News)

monitoring sites for long-term observations, climate forecasting and sea route planning through simulations, development of data archiving systems, providing information to the local community and fostering the next generations of scientists in the field–these are all areas where the Arctic countries could use fresh eyes to do better, and for Team Japan, being an outsider helps," Prof. Enomoto says.

The Team Japan's goals include setting new, higher standards for data gathering.

"We seek collaborations from the U.S. and Canada for simultaneous data-gatherings. We propose new observation methods to improve upon forecasting accuracy. All of this enables us to collect more and better data," Prof. Enomoto says.

Among all parts of the Arctic, Greenland is seeing the fastest climate change, according to Prof. Enomono. Greenland is the world's largest island with 80 percent of the area covered with ice sheet. In other words, Greenland comprises one of the two ice sheets that exist on earth with the other one being Antarctica. The ice thickness ranges from 1,500 meters near the shorelines to 3,000 meters inland. Ice slowly moves from its center to the shore and becomes an iceberg when it falls into the sea. (See Science Report 003)

"Greenland is a forerunner of climate change. Changes occurring there gives us an idea of what may also happen in Antarctica down the road," Prof. Enomoto says. "On the ground, compressed layers of snow become an ice sheet. Lakes that dot the Scandinavian region and Canada are the remnants of an ice sheet from the glacial period that melted. The weight of the ice pushed up the ground to form hills. Greenland's ice sheet is melting fast right now, which could eventually yield a series of lakes when the ice is all gone. At the COP 21, the Conference of the Parties that took place in Paris in 2015, participating countries set the goal of limiting the global temperature increase to a minimum of 1.5 degrees Celsius and a maximum of 2 degrees Celsius above pre-industry levels. Some predict an increase above 1.5 degrees could cause Greenland to melt away rapidly."

Engagement with the local community is essential to the successful operation of research and observation projects. (See Science Report 005)

"We want local people to understand why we are here, and join us in data-gathering. We want them to help us with their local knowledge and experience and show us how to get around. We would share our data and analysis to help improve the quality of the locals' lives. This is a challenge to communicate and build the relationship. We don't want to come across as foreigners from a faraway country who invaded a population's space and set up strange-looking measuring apparatus everywhere," Prof. Enomoto says. "As a first step, we have been hold a series of workshops open to the local residents," he says.

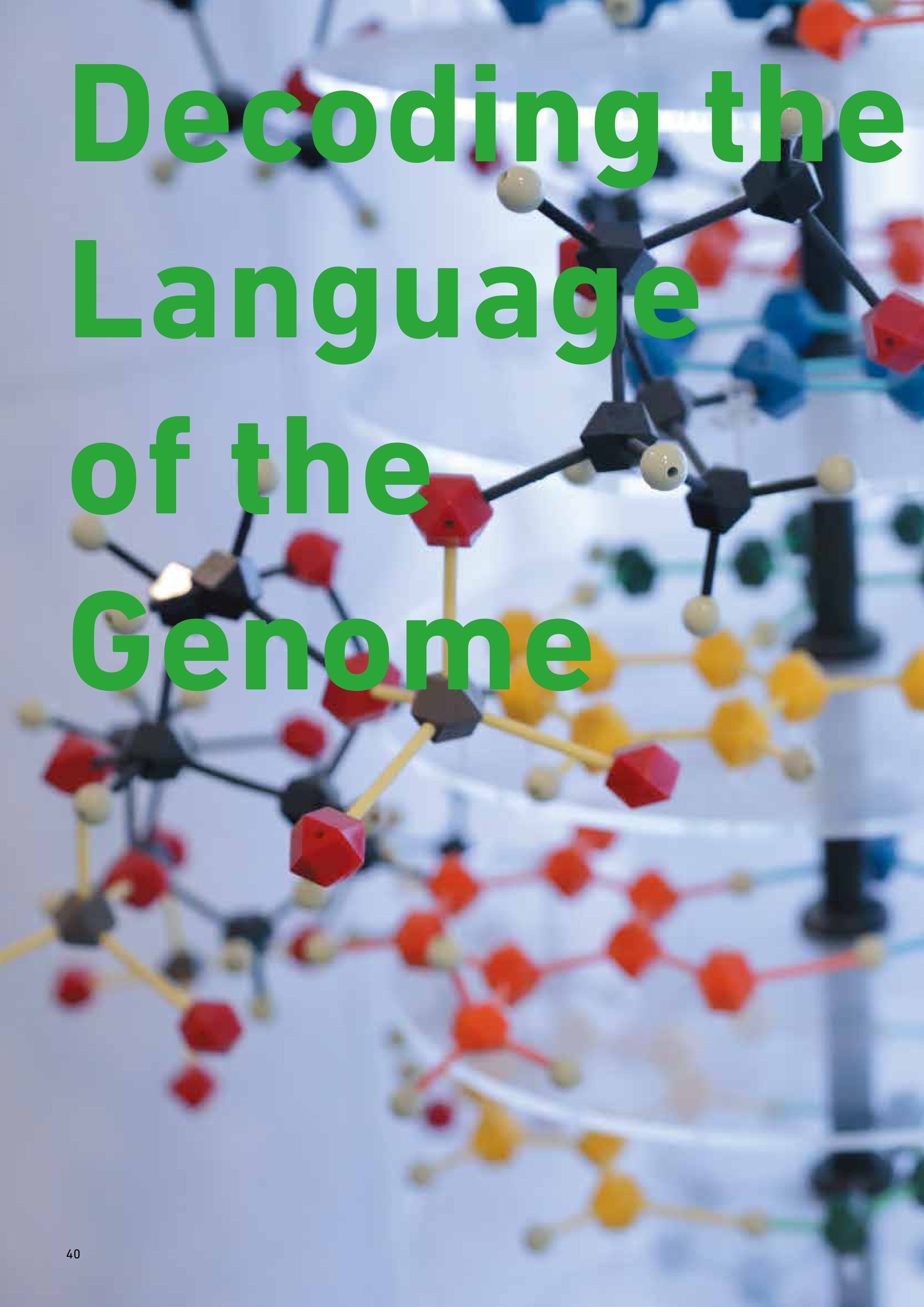
Team Japan has introduced the Arctic Data archive System (ADS), which brings together all the monitoring data, to the locals at one such workshops and was surprised by the hugely positive response of the participating residents.

"The conversation we had with the residents made us realize that we should include more social science data in addition to science data," Prof. Enomoto says.

Prof. Enomoto says Team Japan is striving to provide a wider array of information to foster connections among the ecosystem, industry, global politics and regional security.

Released on: Sept. 11, 2017 (The Japanese version released on April 10, 2017)

Decoding the Language of the Genome



What Have Bacterial Genomes Got to Do with Us?

Ask an Expert: Ken Kurokawa
Professor of National Institute of Genetics (NIG)

Dr. Ken Kurokawa is professor at the National Institute of Genetics (NIG) and the director of the NIG's Advanced Genomics Center. Since 2014, he has serviced as a director of "Hadean Bioscience," a project funded by the Grant-in-Aid for Scientific Research on Innovative Areas . Dr. Kurokawa, who studied geology and tectonics in the undergraduate and master's programs, holds a bachelor's degree from Tohoku University and a Ph.D. from Osaka University. He joined the NIG in 2016 after serving as associate professor at the Nara Institute of Science and Technology and as professor at the Tokyo Institute of Technology. After founding Japan's first full-scale metagenomic project in 2004, Dr. Kurokawa published the analysis of the human intestinal microbiota from 13 Japanese individuals. Dr. Kurokawa is known for developing the metagenomic analysis technique he used for the project to study how microorganisms in the soil and hot spring waters change in relation to the environment.

In 2003 when the world celebrated the completion of human genome mapping, Craig Venter, an American geneticist who was instrumental to the project's success, had already moved onto his new target: Marine bacterial genomes.

He stunned the public in April 2004 with the announcement that he and his team discovered 1.2 million new genes following their research voyage around the Sargasso Sea. The news signaled a new era of metagenomics–genetic analysis of microorganism samples directly taken from a particular location to gain a broad understanding about the environment. Metagenomics differs from the traditional genetic research in that it puts microbes found in the wild straight through a sequencing machine instead of using microbes cultured in laboratories.

Because microbes have direct control over both our bodies and environments, decoding microbial genomes and their relationships to the surroundings can pave the way for so many scientific breakthroughs from the medical to pharmaceutical fields. Some people are already anticipating agricultural, construction and various other applications of the information that metagenomics can provide.

With the speed and the scale of genetic analysis increasing ever more, thanks to the emergence of next-generation sequence technologies, how can we expect to benefit from metagenomics? --- A leading expert from the National Institute of Genetics will explain.

What Is A Genome Anyway?

A genome refers to the entire set of an organism's genetic information, and consists of deoxyribonucleic acid (DNA). DNA is comprised of four nitrogenous bases known as AGCT (adenine, guanine, cytosine and thymine). AGCT are attached to DNA's double helix in a sequence unique to each species. The human genome contains about 3 billion bases. Candidatus Carsonella ruddii, a parasitic microorganism, has the smallest genome discovered to this day with 159,662 bases, which is equivalent of the number of letters contained in a typical Japanese 24-page broadsheet morning newspaper.

A codon, or a genetic code, is a sequence of three bases. Each codon can attract a specific amino acid, and its base sequence order dictates which one of the 20 amino acids it can attract. Codon patterns are transcribed from DNA into messenger RNA (mRNA), which are then translated to create a chain of matching amino acids for the production of protein. This DNA-to-protein flow of genetic information is universal among all species.

"Yes, that's all species–including us humans," Prof. Ken Kurokawa of the National Institute of Genetics says, "Genome is a bridge between data and its physical manifestation."

The Tama River system supplies about 20 percent of Tokyo's drinking water. Photographed here is a downstream section of the river where Prof. Kurokawa and his students conducted research. Prof. Kurokawa credits the students for the successful analysis of the riverfront microbe genomes.

The Hakuba Hoppo Hot Spring in Hakuba Village in Nagano Prefecture gushes out through sepentinite, featuring 50 degrees Celsius (120 degrees Fahrenheit) water with high hydrogen and alkali (pH11), content. This hot spring environment is believed to run parallel to that of the hadean Earth. A group of scientists is conducting metagenomic research on the microbial community that thrives in the boiling water bubbling 600 meters (1,968.5 feet) underground inside of the sepentinite. The work is beginning to shed light on the primitive metabolic system.

Microorganisms as a Dependable Research Tool

The most fascinating aspect of the genome, Prof. Kurokawa says, is that it lends an objective yardstick to define all species on equal terms. In genomic studies, all living matter from humans to microbes boil down to A, T, G and C.

And in the metagenomics, "those four chemicals can add infinite layers of meaning to the environmental context that you are trying to understand," Prof. Kurokawa says.

For example, people will know what conditions their digestive systems are in by analyzing their human intestinal metagenome.

"We could tell you what the temperature inside your stomach is by just looking at the genomic analysis of your intestinal metagenome without ever taking the body temperature," says Prof. Kurokawa. "The analysis can also show you the pH value (hydrogen-ion exponent) and humidity levels of the environment, which helps you get a good handle on the sanitary condition of your surroundings. On top of that, if you take time series data, you will be able to see how the environment has changed over time, as well," Prof. Kurokawa says.

In short, microbial genomes contain a treasure trove of information useful for gaining insights into the world around us. Metagenomics attempts to deliver that information to us through multidimensional analysis and interpretation of data.

But how can microorganisms offer this much data to decipher?

"That's because microbes always respond to environmental factors–the sole controlling factors in changes to organisms–in the same, specific manners," says Prof. Kurokawa. "You can always count on microbes to provide the most objective data. Metagenomes just have no room for human manipulations."

For the same reason human microbiota can help determine the condition of digestive systems, microorganisms can contribute to soil research.

"We conducted research on the waterfront microbial communities along the entire 138-kilometer Tama River from the Okutama Lake in the western edge of Tokyo through the Haneda Airport on the Tokyo Bay. We then integrated socio-environmental data into it, such as population distributions and crime rates of the areas. This provides us a comprehensive picture on the community environment. We are now working to develop technology that enables us to assess a neighborhood environment simply from taking a look at its microbial community structure," Prof. Kurokawa says.

This illustrates how metagenomics can layer countless individual data sets covering ranging topics to generate multifaceted yet cohesive information, Prof. Kurokawa says.

"By using the genome as an equal gauge for all living matter, you can create linkages among various research fields and industries," Prof. Kurokawa says. "And, that's the crux of the premise of metagenomics," he says.

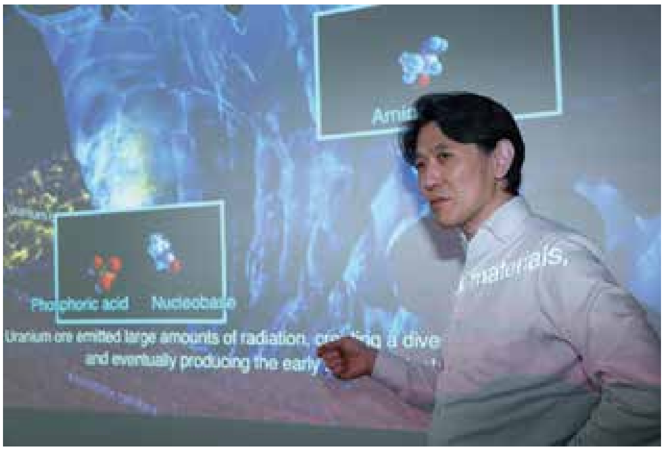
Human and Microbial Genomes: We Are All in This Together

Success of metagenomics depends on the sizes and scopes of accessible databases. There are currently 170,000 samples of microbiome available to the public, and Prof. Kurokawa's lab is building a "microbe GPS" that shows the environmental contexts for the organisms when the data is plugged into it.

"We already have catalogs of microbes from all corners of the globe. By layering other types of information to the data, we should be able to figure out to a reasonable degree what the presence of certain microbes means for the environment," Prof. Kurokawa says.

He believes the "hologenome" approach that includes human genomes into the mix for data analysis is also a powerful method for understanding the environment.

"Our human cells and the microbes are all made from the genomes, and they cohabitate in the same environment. The



The Hadean Bioscience Project, in which Prof. is involved as the NIG's representative, has many images and videos that explains origin of life with computer-generated graphics and animations. For "The Whole History of the Earth and Life" series on YouTube, please go to https://www.youtube.com/channel/UCCToEEPlF7ur1m0efQTHG4g.

idea of treating all the genomes and environmental data as part of one big system and decoding the relationships among them is growing into this new approach," Prof. Kurokawa says.

Understanding the Earth Environment t Understand the Origin of Life

Prof. Kurokawa has always approached genetic studies with a simple mantra: To understand life, one must understand its environment.

With that same principle, Prof. Kurokawa is now tackling the origin of life study, envisioning the hadean Earth between 4.6 billion-4 billion years ago.

"The reason life scientists can't begin to fathom how life began on Earth is because we don't know anything about the Earth environment of that time," Prof. Kurokawa says.

So, some leading scientists in geology, earth science, life science and planetary science decided to get together and discussed the matter.

"We found out that hadean Earth surface environment was inhospitable to life. Seawater was highly acidic with a pH value of zero, and melted rocks were swirling in it, like iron inside a blast furnace. Forget the life-nurturing image that we have of the oceans today. Even if life was born, it wouldn't have survived too long in the harsh environment," Prof. Kurokawa says. "The scientific evidence collected thus far suggests that life may have emerged amid the process of ferocious chemical reactions triggered by radioactivity inside a geyser-like spring on a continent in the primitive era."

But, Prof. Kurokawa points out that scientists still have a long way to go to decode how life began.

"How the genome became responsible for carrying self-replicating information remains a mystery," Prof. Kurokawa says.

The Advanced Genomics Center: The Axis of International Academic Collaborations for Genomic Analyses

The Advanced Genomics Center is Japan's central institution for genomic studies. As one of the country's 19 inter-university research organizations, the center provides universities and researchers with access to databases containing voluminous genome data compiled with the use of next-generation sequencing technologies. The center promotes resource-sharing and joint research among all genome research institutions and also undertakes large-scale sequencing and analysis projects in response to requests by universities. The center opened in October 2011 in Mishima City in Shizuoka prefecture and has since conducted various cutting-edge genome research in collaboration with other relevant organizations.

The Advanced Genomics Center
http://genome.nig.ac.jp/

Released on: Nov. 10, 2017 (The Japanese version released on July 10, 2017)

The Next Medical Revolution Could Come from Innovation in DNA Sequencing Technology

Ask an Expert: Yutaka Suzuki
Professor of Department of Computational Biology and Medical Sciences, University of Tokyo

Dr. Suzuki is a professor at the Graduate School of Frontier Science at the University of Tokyo. He earned his bachelor's degree in 1994 and his doctoral degree in 1999, both from the University of Tokyo. After working for RIKEN (Rikagaku Kenkyujo)'s Genome Science Center as research associate in 1999, he moved to the Laboratory of DNA Information Analysis at the Institute of Medical Science at the University of Tokyo in 2000 to serve as assistant professor and became associate professor in 2004. For the past 25 years, Dr. Suzuki has conducted research focused on transcriptome collaborated with his mentor, Prof. Sumio Sugano of the University of Tokyo. His research is aimed at gaining a systemic understanding of how genomic mutation manifests in the form of cancer, neurologic disorders and other types of diseases.

The genome is a set of nucleotide sequences that makes you who you are and different from everybody else.

Determining what your sequence looks like used to require enormous time and money. But, thanks to the immense advancement of sequencing technology since the first successful human genome sequencing nearly 15 years ago, we can now use so-called next-generation sequencers to analyze our genomes far more quickly, accurately and cost effectively than ever possible before.

With information on an astronomical number of genomes stored in databases, the big data is helping further advance genome science. It is also revolutionizing medicine and enabling improvements in multiple aspects of our lives from the environment to land and economic development.

With sequencing technology continuing to progress at a rapid pace, it promises to transform our ourselves. So, where is this all headed? Dr. Yutaka Suzuki, a human genome expert, will explain.

Delivering Cutting-edge Medical Treatment Anywhere, Including the Middle of a Jungle

A handful companies that developed the next-generation sequencing technologies control the entire global sequencing market. This means that the base technologies that support the industry all come from these same companies. The monopoly translates into uniform research procedures, including how to verify experimental data and conduct follow-up experiments, says Prof. Suzuki.

"Having said that, next-generation sequencers present tremendous possibilities for further advancement and revo-

lutionizing medical practice. We must decide how we want to benefit from the technology and put the research community on the right path toward that goal," Prof. Suzuki says.

As a step to do just that, Prof. Suzuki has launched the "on-site genomic diagnosis" project to provide better medical access to people in undeveloped countries. The project's goal is to develop a compact, portable sequencer that can be used anywhere to quickly identify the best course of treatment for patients and deliver it on the spot. One scenario would involve drawing blood from a patient suffering a high fever in the middle of a jungle and conducting high-speed sequencing of the patient's genome conducted right on the site. The result would reveal not only what's causing the fever, such as malaria and dengue, but also what strain or type of virus or parasite it is, whether it would respond to medication and what other underlying infections the patient may have. One requirement for the sequencer, in addition to its portability and sequencing speed, is the ability to withstand tropical heat and elements, and Prof. Suzuki and his team are working to develop a group of technologies that would enable the use of sequencers in non-hospital, non-sterilized environments, including outdoors.

"These technologies have a surprisingly wide range of applications, including self-health check through genome sequencing of your own stool sample taken at home, discovering a new species during field research, verifying the origins of seafood like eel and other food products to prevent false claims and identifying individuals involved in a crime through genome sequencing of hair," Prof. Suzuki says.

"Genomic diagnosis" could particularly be useful for quarantine checks at the airport, he says.

Using Genomic Mutation as Weapons Against Cancer

When it comes to medical use of genomic information, whole genome sequencing is only the first step. Many researchers are now focused on finding out how someone's DNA molecules are organized as a system in order to use that information to predict the person's medical risks, or keep a disease from progressing.

The prototype sequencer for genomic diagnosis is small enough to fit in your palm. The sequencer that Prof. Suzuki is shown holding here needs to be inserted into the specimen holder, which allows the user to read the processed data and analysis on one's personal computer.

Single-cell sequencer that deciphers and provides genomic information of each separate cell.

Prof. Suzuki holds a tray for a single-cell isolator/analysis machine. Trays are divided into microscopic sections that function as receptacles for separated cells, allowing the machine to read genomic information of each cell.

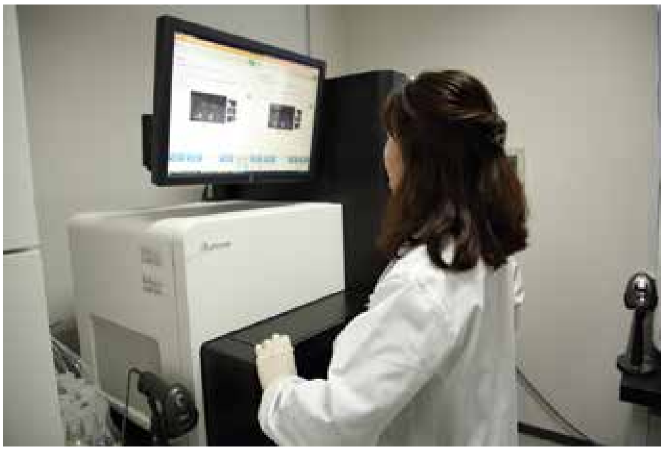
Each living organism has just one set of DNA information, but mRNA that transcribes information from DNA for synthesizing protein can alter in response to various factors and can carry different information in different organs. Prof. Suzuki, who has studied mRNA transcriptome for many years, says analyzing gene readouts is useful in cancer diagnosis and treatment.

“Cancer is a disease that causes a transcriptome mutation. We can find out how a genome mutation affects gene expressions and what happens to it if we try to correct it with medication,” Prof. Suzuki says. “Causes and prognosis largely vary from patient to a patient because of the genome. This explains why there are 500 to 1,000 different types of lung cancers alone. Just because a medication worked for one patient, that doesn’t mean it will help another patient. Everyone has their own story, and we hope genomic mutation information will help us deliver individualized treatment.”

The Tama River system supplies about 20 percent of Tokyo’s drinking water. Photographed here is a downstream section of the river where Prof. Kurokawa and his students conducted research. Prof. Kurokawa credits the students for the successful analysis of the riverfront microbe genomes.

Taking Out-of-the-box Approach to Expedite Drug Development

The Suzuki Lab is also working to develop methods for integrating sequencing results and prognosis obtained through computer simulations in order to make individualized



A researcher operates a next-generation sequencer. A genome sample needs to be placed in the black box on which the researcher has her hands. The user can control the machine via the monitor and get sequencing information.

genomic treatment a reality. Chemotherapy may kill 99 percent of cancerous cells, but the remaining 1 percent that doesn't respond to the drugs often triggers a recurrence.

"It's crucial to identify the 1 percent cell with high accuracy. We then need to perform genomic analysis of that cell and use the result for simulations," Prof. Suzuki says.

In the meantime, "immune checkpoint inhibitors," a new type of cancer medication that activates the patient's immune system to attack cancerous cells, has drawn public attention in recent years. Prof. Suzuki points out that the ability to distinguish and isolate cells to target is important in this type of treatment, as well.

Single-cell Technology x Next-generation Sequencers for Cutting-edge Cancer Treatment

Single-cell technology refers to the technology for isolating and analyzing a single cell. When trying to analyze a genome sample with a next-generation sequencer, researchers normally mush cells before putting them through the machine. In case of cancer, however, they aren't trying to evaluate the sample as a whole.

"In partnership with Dr. Ryoji Yao, principal investigator of the Division of Cell Biology, Cancer Institute, The Japanese Foundation for Cancer Research, we are conducting research on single-cell technology for isolating each cell and analyzing it with next-generation sequencers. Our initial focus is colorectal cancer," Prof. Suzuki says.

The project is part of the "Platform for Advanced Genome Science" that uses revenues from genome sequencing to augment grants that support researchers. "We have high hopes that the project will elucidate how colorectal cancer cells are structured and particularly how would-be cancerous cells emerge and mutate," Prof. Suzuki says.

Platform for Advanced Genome Science: Support Network for Genome Analysis Research led by the Inter-University Research Institute

"Platform for Advanced Genome Science (PAGS)" is a project of KAKENHI, the Ministry of Education, Culture, Sports, Science and Technology (MEXT)'s Grants-in-Aid for Scientific Research program. PAGS provides cutting-edge genome sequencing and analysis technology to KAKENHI supported projects. The National Institute of Genetics (NIG), which is located in Mishima, Shizuoka Prefecture and is one of the Inter-University Research Institutes, serves as the central institution for PAGS in which various universities and research organizations participate to form a network. PAGS has helped various research projects generate multidisciplinary research outcomes to promote cutting-edge life science findings, as well as the advancement new genome technologies.

Platform for Advanced Genome Science (PAGS)
https://www.genome-sci.jp/e

Released on: Dec. 11, 2017 (The Japanese version released on July 10, 2017)

How Will Genomic Discoveries Change Our World Views?

Ask an Expert: Yuji Kohara
Specially-appointed Professor of National Institute of Genetics (NIG)

Dr. Kohara is a specially appointed professor at the National Institute of Genetics(NIG) and the director of the ROIS's Database Center for Life Science(DBCLS). He serves as the principal investigator of "Platform for Advanced Genome Science(PAGS)", which is one of the Ministry of Education, Culture, Sports, Science and Technology(MEXT)'s Grants-in-Aid for Scientific Research program. After receiving a PhD's degree in science from Nagoya University, he worked as a research fellow for Medical Research Council's Laboratory of Molecular Biology in the United Kingdom. He joined the NIG as professor in 1996 and served as its director from 2004 through 2012. Amid the fast advancement of molecular biology, Dr. Kohara has worked to decode the connections between the genome and phenotypes. As data gathering and sharing are key to promoting genetic research, Dr. Kohara is also focusing on the development of integrated databases.

Thanks to the advancement of technology, we now sequence countless genomes every day, including personal samples and the DNA of living matter around us. All such data becomes aggregated to form big data, which gives us new insights into life. For example, we now know what separates humans from chimpanzees is a 1.23 percent difference in nucleotide sequences of the two species' DNA. Big data also led to some revelations that resulted in a rewriting of the phylogenetic tree of evolution.

The genome provides scientific evidence useful to prove or disprove hypotheses. The demand for this versatile research tool continues to grow in research fields, many of which have had little to do with the genome until recently.

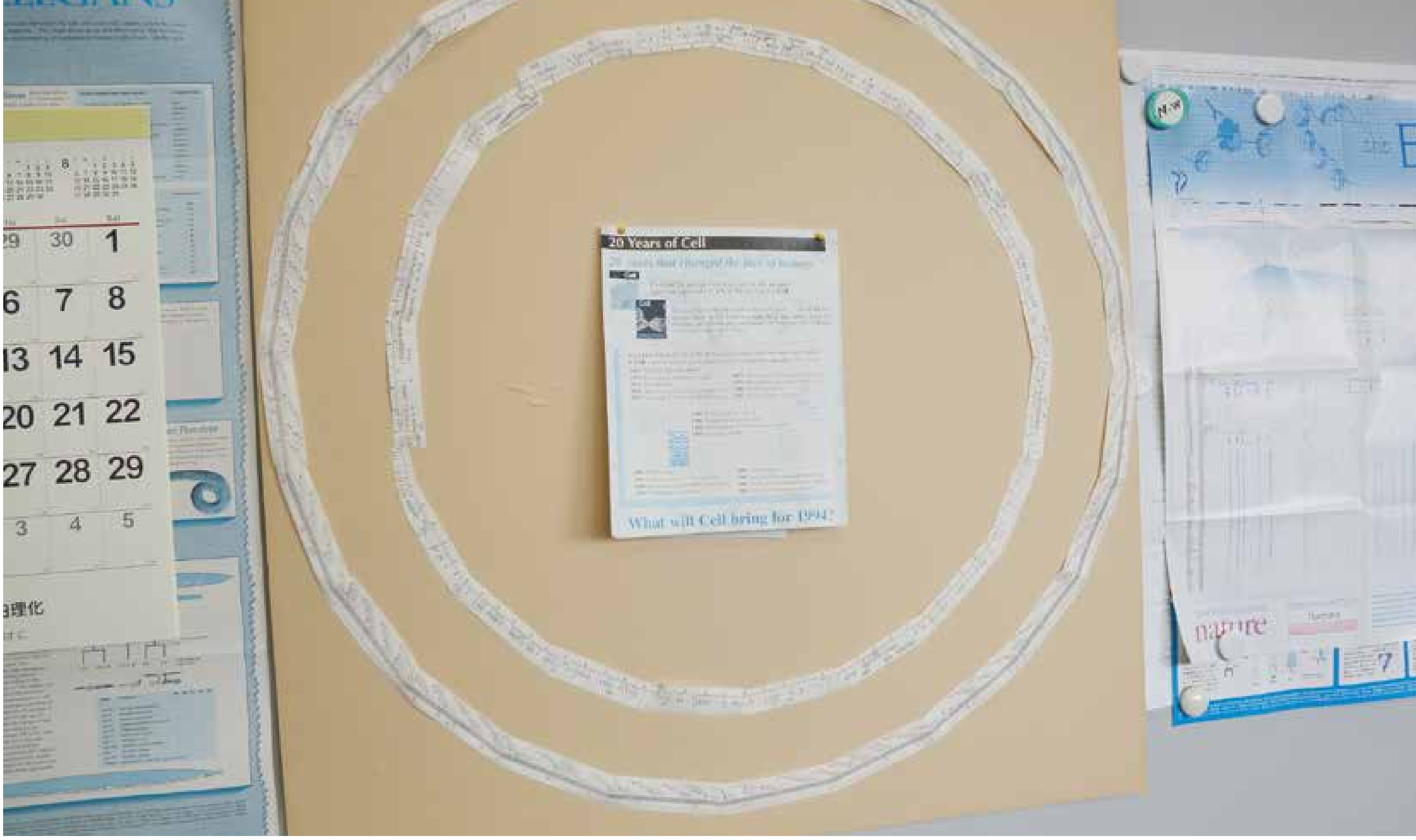
But what does this really mean for our future? How are researchers trying to leverage the technology? And how will our new discoveries change our perspective on life?

Dr. Yuji Kohara of the Database Center for Life Science will share his view.

One Technological Breakthroughs After Another

Genomics is a fiercely competitive research field, where not only scientists but also technologists from around the world are racing to innovate. To put the speed of evolution in perspective, "a couple of major technological innovations emerge every five years," according to Prof. Yuji Kohara of the National Institute of Genetics, who serves as the director of the ROIS's Database Center for Life Science (DBCLS).

The industry-standard genome sequencer also needs to be



replaced with a newer version every so often to keep up with technological advancements.

"As soon as a new machine becomes available, we must adopt it and adjust our procedures and operations accordingly," Prof. Kohara says.

In other words, research labs would need to review and modify their entire research process, from what types of experiments to carry out to what samples to prepare to how to conduct analysis. Such reconfiguration requires high levels of expertise each step of the way.

In the meantime, genomic analysis as a research tool continues to make its way into various research fields.

"Researchers who weren't even thinking about the word, 'genomes,' yesterday are suddenly talking about sequencing them today. We are seeing more and more of this," Prof. Kohara says. "This trend is particularly noticeable in the environmental, medical and agricultural fields. The genomic analysis of river soils and bacteria, for example, can reveal interesting facts. Medical researchers are now interested in doing not only genomic diagnosis but also genomic analysis of intestinal bacteria. Agricultural researchers are looking to use genomic analysis for potential crop improvements."

The demand for single-cell genomic analysis is also on a rise, Prof. Kohara says.

Individual labs only cannot do so much to catch up with rapid technological changes by themselves. But, what if there was a centralized system that pulls together all the expertise needed for making other important changes–such as software and hardware upgrade–and makes it accessible to labs?

This idea led to the launching of "Platform for Advanced Genome Science (PAGS)." PAGS is a project of the Ministry of

Genomic analysis of river soil and bacteria is particularly popular.

Education, Culture, Sports, Science and Technology (MEXT)'s Grants-in-Aid for Scientific Research Program known as the KAKENHI program. PAGS makes cutting-edge technologies for genome and data analysis available to researchers when requested in order to help them carry out their KAKENHI research projects.

In 2000, a core-facility team was formed within a KAKENHI project called Grant-in-Aid for Scientific Research on Priority Areas. This team was the precursor to PAGS's research support team, according to Prof. Kohara. This team's primary focus changed over time to aiding genomic research. It also redesigned its structure to facilitate multi-disciplinary distribution of resources and technologies to become the NIG-led support system that it is today.

PAGS's research support team coordinates resources and expertise that university labs request for their KAKENHI projects. The team assists universities with the alignment and maintenance of the latest equipment and technologies and conducts genomic analysis for some of the most advanced research projects. Prof. Kohara represents the PAGS.

"As PAGS's support team adopt newer equipment, it is becoming capable of sequencing genomes faster and more accurately than ever. Informatic scientists in PAGS's support team have access to more actual data through assisting university labs, which, in turn, is helping them improve their analysis skills," Prof. Kohara says. "PAGS's support team also coordinates a one-week on-the-job training program for students to gain necessary skills and better contribute to the KAKENHI projects for which they work. In other words, our team helps put KAKENHI labs and PAGS' support team on a path for continuous improvements."

If You Want to Win the Competition, Build Databases

Genomic data is critical to life science research and becomes more so as data accumulates.

"In the game of big data analytics for genomic research, the one with the most amount of data wins," Prof. Kohara says. "In addition to the U.S., places like China have mega-analytics centers. These are the places that get the largest amount of data because data naturally goes to analytic service providers. This explains why Japan needs to come up with a way to attract and compile more data, so Japanese life scientists have more data with which to work," Prof. Kohara says.

But, in order to truly tap into the potential of genomic big data, one must have not only a substantial amount of data but also information that puts the data in context.

"For example, in order to use genomic big data for medical research, you need to link genomic data with epigenetics or medical records to glean knowledge from it," Prof. Kohara says. "Although 'we are all made from genomes', all those bacteria and other life matter that live with us help shape us," Prof. Kohara says.

Hologenome refers to sampling and analyzing of genomes in the entire environment that surrounds us, including bacteria outside our bodies and those inside our organs, such as

National BioResource Project, the Database for All Organisms from Prokaryote to Humans

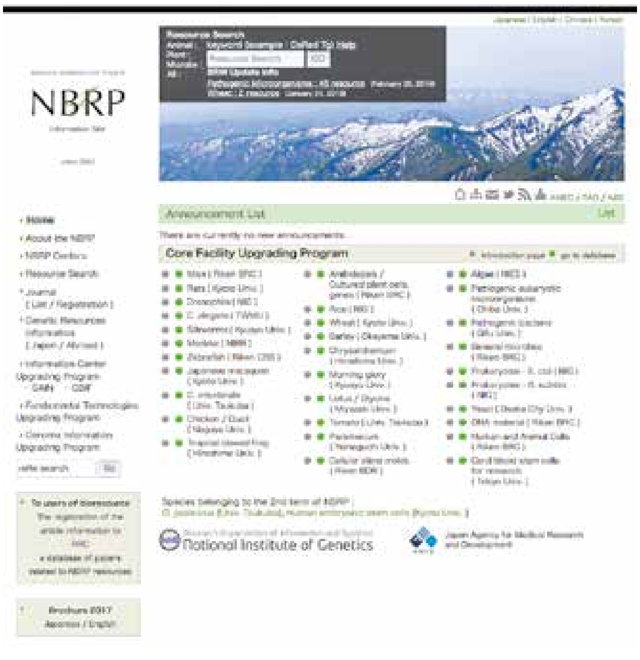
The purpose of National BioResource Project (NBRP) is to collect, preserve and share samples of 30 different bioresources that are regarded as strategically important for Japan's research projects.

Bio-resource refers to "systems, groups, tissues, cells, and genetic materials of animals, plants, and microbes and their information as research and development materials" according to the NBRP website.

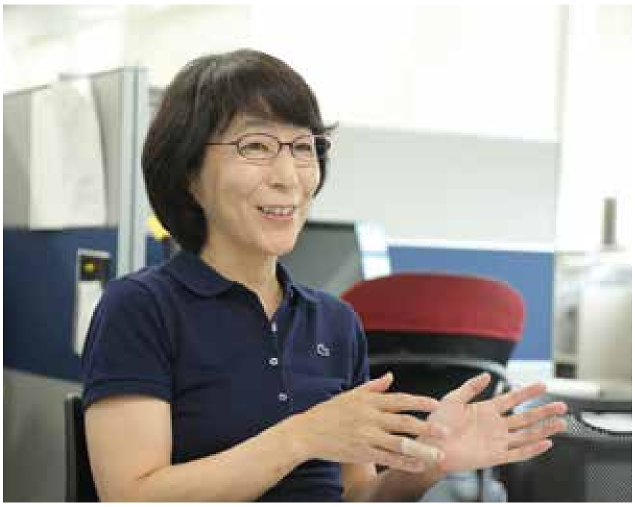
The Genetic Strains Research Center at the National Institute of Genetics (NIG) serves as the NBRP's portal for information and model species. Dr. Shoko Kawamoto, associate professor at the Genetic Strains Research Center, was recently appointed to lead the NBRP.

"We are trying to build a bio-resource database that allows for cross-species search," Dr. Kawamoto says. "That means the database has to have sequencing data of all model species' genomes as well as analytics on the bulk of those genomes."

Once common denominators among different species become clear, it enables replication of a scientific experiment with multiple species. For example, when a newly developed variety of wheat grows well, one might be able to create a same type of variety of another monocot.



National BioResource Project (NBRP) Official Website
http://nbrp.jp/
More than 28,000 research papers have used NBRP resources.



Dr. Shoko Kawamoto, associate professor at the Genetic Strains Research Center, was recently appointed to lead the NBRP.

"It might also allow replacing mice and rats with their counterparts that are structurally much simpler, such as zebrafish and drosophila, for experiments in research on human diseases. There is a growing expectation that this would help speed up research on rare diseases,"Dr. Kawamoto says. "The NBRP could potentially match clinical researchers with drosophila researchers," she says.

In addition, the NBRP serves as a Japanese arm of the Global Biodiversity Information Facility (GBIF), an "open-data research infrastructure" that gathers and provides access to information about all kinds of species to better understand ecosystems around the world. The NBRP has a biodiversity registry through which people can report what species lives and where.

"Research data on spiders on the island of Yakushima and swallows nestling in the Kinki region are examples of what we've collected through the registry," Dr. Kawamoto says.

But having an intelligent database may not help as much if researchers don't know how to tap into it, says Dr. Kawamoto, who has the experience as a genetic researcher.

"Demand for personalized medical treatments is expected to grow. We are going to keep that in mind to build an open science type of database system that caters to the needs of end users like patients," Dr. Kawamoto says.

the stomach.

As hologenome research continues to advance, the line between the human and the environment may become more ambiguous. We may not be just the bodies that are covered with the skin after all; things around us could also be considered as a part of us.

"That can also lead to the profound question of what life is and what you consider as an individual."

Released on: Jan. 25, 2018 (The Japanese version released on Aug. 10, 2017)

The Making of the Databases of Scientists' Dream

Ask an Expert: Susumu Goto
Professor of Database Center for Life Science (DBCLS)

Dr. Susumu Goto (right) is a professor for Database Center for Life Science (DBCLS) at the Research Organization of Information and Systems' Joint Support-Center for Data Science Research. Before joining the DBCLS in 2017, he worked for 23 years at Kyoto University Institute for Chemical Research, where he helped Prof. Minoru Kanehisa found a bioinformatics database called Kyoto Encyclopedia of Genes and Genomes (KEGG) and developed his expertise in bioinformatics, an interdisciplinary science field that uses computer to sort and analyze big biological data. Dr. Goto earned his bachelor's and doctor's degrees in computer science in 1989 and 1994 respectively, both from Kyushu University. Dr. Toshiaki Katayama (left), a Project Assistant Professor at the DBCLS, has led BioHackathon, the DBCLS's international coding camp for life science data standardization, since its inception 10 years ago.

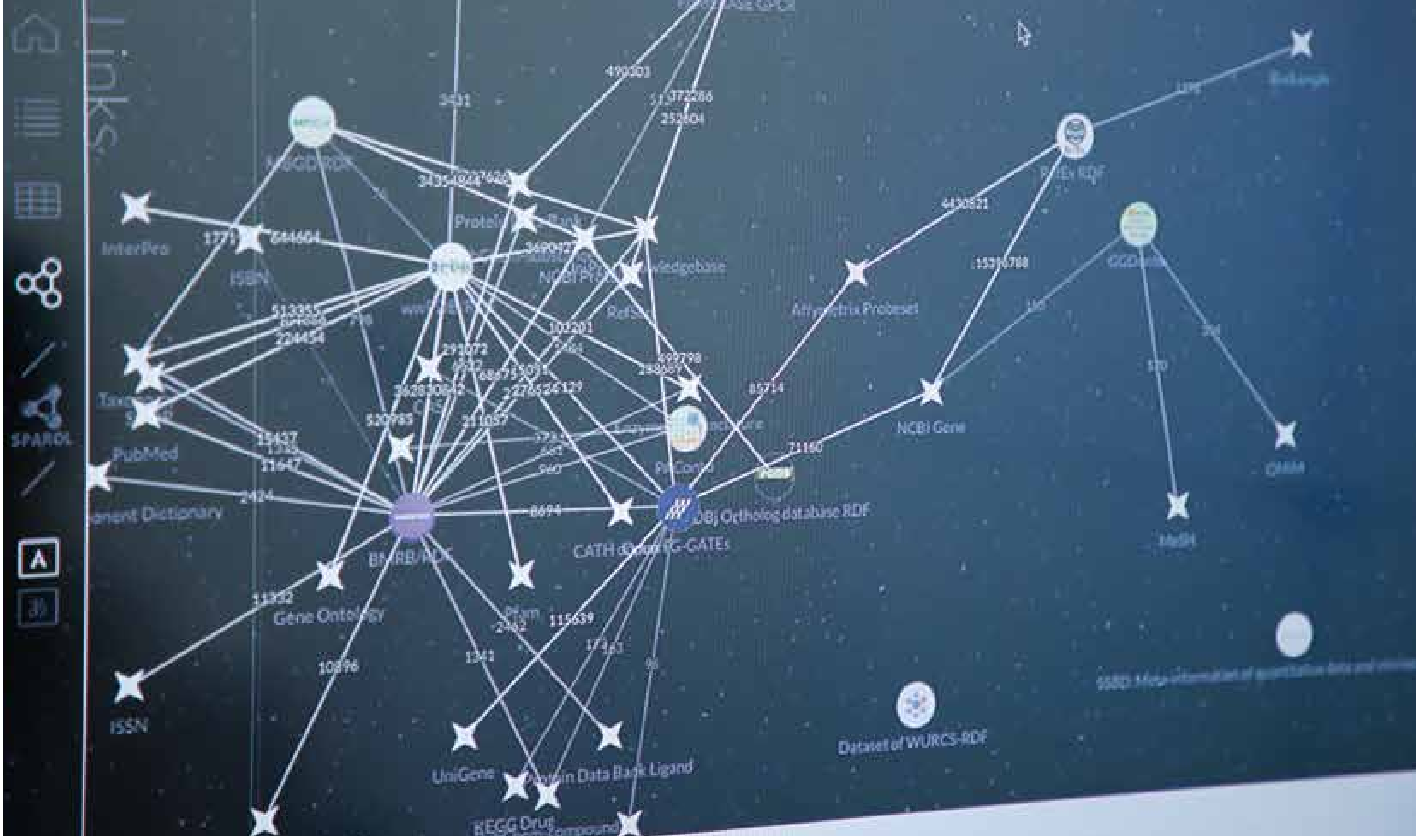
Isaac Newton experienced his epiphany while in a garden with apple trees, and Archimedes, a bathhouse. In the case of modern life scientists, their databases could become their go-to place to seek inspiration to find their own eureka moments.

Bioinformatics experts from around the world, including those in Japan, are making progress on the decades-old initiative to develop ways to blend all sorts of life science data and create the most comprehensive and cohesive reservoir of information imaginable to accelerate scientific discoveries. The ultimate database they envision would not only make an astronomical volume of information available at the user's fingertips, but also show how seemingly unrelated data are connected in the real world. For example, data on a human gene could encompass anything from the explanations about how the same gene works in other species' systems to what diseases are associated with the gene.

With big data growing even bigger and more important in life science, there is an impetus for the movement to consolidate scattered biological data. Two experts from the Database Center for Life Science (DBCLS) who have been on the forefront of the movement will explain what progress has been made thus far in Japan and what challenges still lie ahead.

Database as an Encyclopedia of Biology

Life scientists began dreaming up integrated biological databases during the 1990s as they became flooded with billions of pieces of genomic data. Among those dreamers were scientists at Kyoto University Institute for Chemical Research, who envisioned creating a virtual encyclopedia of biology.



They appropriately named their project, Kyoto Encyclopedia of Genes and Genomes (KEGG).

"Back then, all you saw in databases was character data without context, such as nucleotide acid and amino acid sequences, whereas what's in textbooks was mostly graphical data, explaining how metabolic pathways and signal transduction pathways work," Prof. Goto, one of KEGG's founders, said, looking back on the early days. "We thought it would be great if we could combine both types of information. We figured the web, which was becoming popular at the time, would work well as media for that."

Understanding Human Genes in Relation to Other Species' Genes

KEGG is a database that brings together genomic and genetic data with the information about genes' roles in "pathways," such as metabolic pathways and signal transduction pathways.

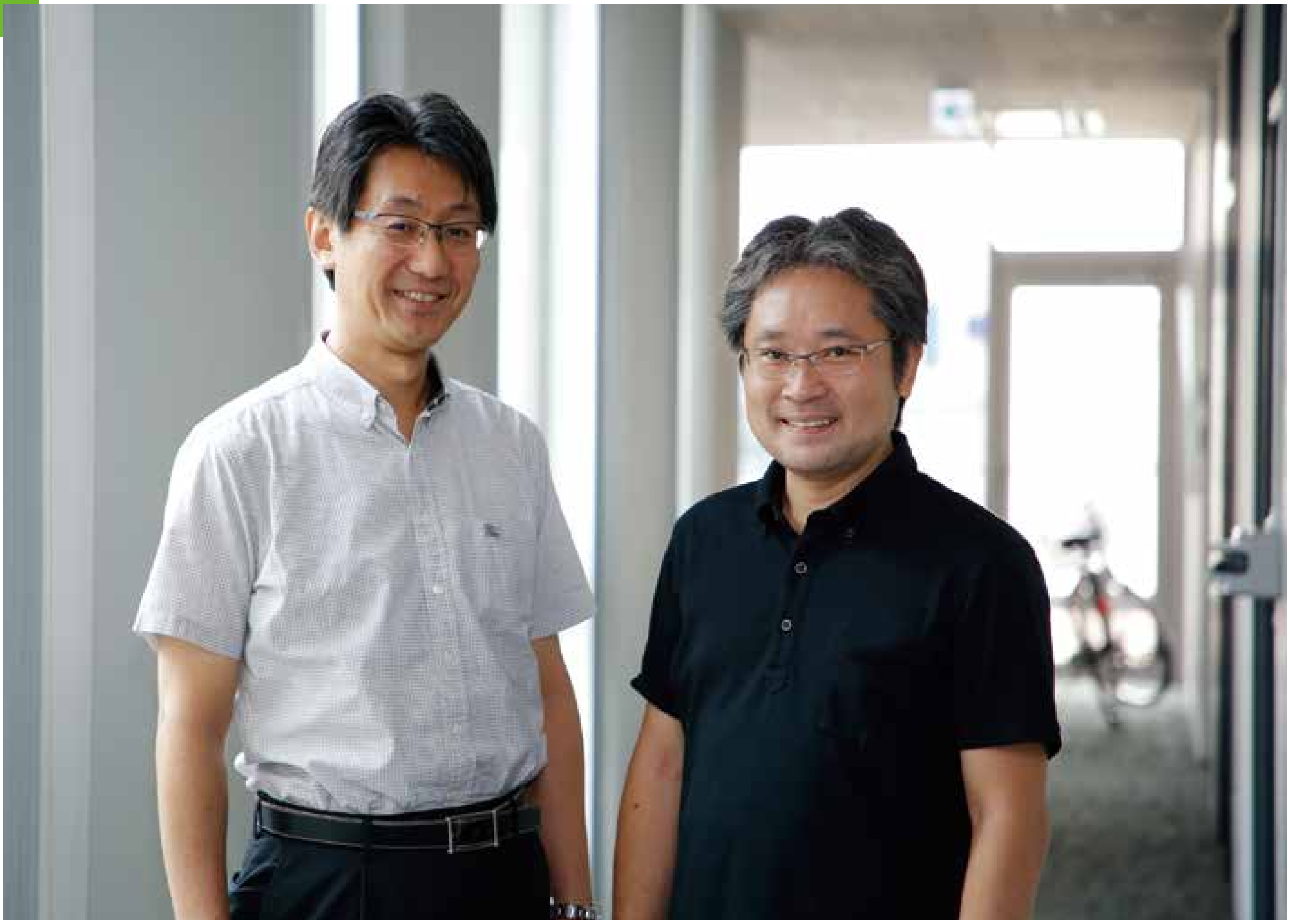
"Once the genomes of the human and tens of thousands of other species were sequenced, it became clear what genes are there. And we can use diagrams to show things like what each gene does to cause a chain of metabolic reactions. KEGG calls the flowchart describing this network a 'pathway map,'" Prof. Goto said.

Pathway maps make it easier for people to grasp and compare the pathway patterns and the genes involved among various species.

"KEGG also catalogue genes found in ocean and soil samples and compile information about the gene functions. So, we expect the KEGG database will enable us to look at the relationships among species from the whole new viewpoint of how one species' chemical processes influence another's," Prof. Goto said.

Standardizing Data to Make It Sharable Among All Databases

There are many databases in Japan that are open to the public. These databases come in all sizes and forms, covering a range of species from microbes to eukaryotic to primates to



birds. Major databases include Tohoku University's Tohoku Medical Megabank that collects human genomic information and Kazusa DNA Research Institute, which, for the most part, deals with plant genomes.

The DBCLS wants researchers to be able to simultaneously access many of these databases and effortlessly cross-reference all kinds of data, including chemical compounds and structures, pathway information and genome data on various species. This means all data must be standardized for integration, and developing technology for it is the key, Prof. Goto said.

"We are using semantic web technology to develop a data integration system that works with a data model called Resource Description Framework (RDF). RDF allows the system to learn the context of data and link data accordingly. So, you can search a gene by an associated disease and vice versa, for example. You can also look something up by gene-disease correlation types," Prof. Goto said. "With this system, it's super easy to locate the information that you are looking for. The system also helps you analyze and interpret data more accurately and efficiently."

While the system is being constantly improved and updated, researchers across Japan can already take advantage of it to obtain diverse data from multiple databases. It's made an open source system so people can use it as it suits them, according to Prof. Goto. "The user might be a bioinformatician interested in using the system's API to develop a proprietary application, or could be a data scientist who wants to extract information from the databases to make a discovery," Prof. Goto said.

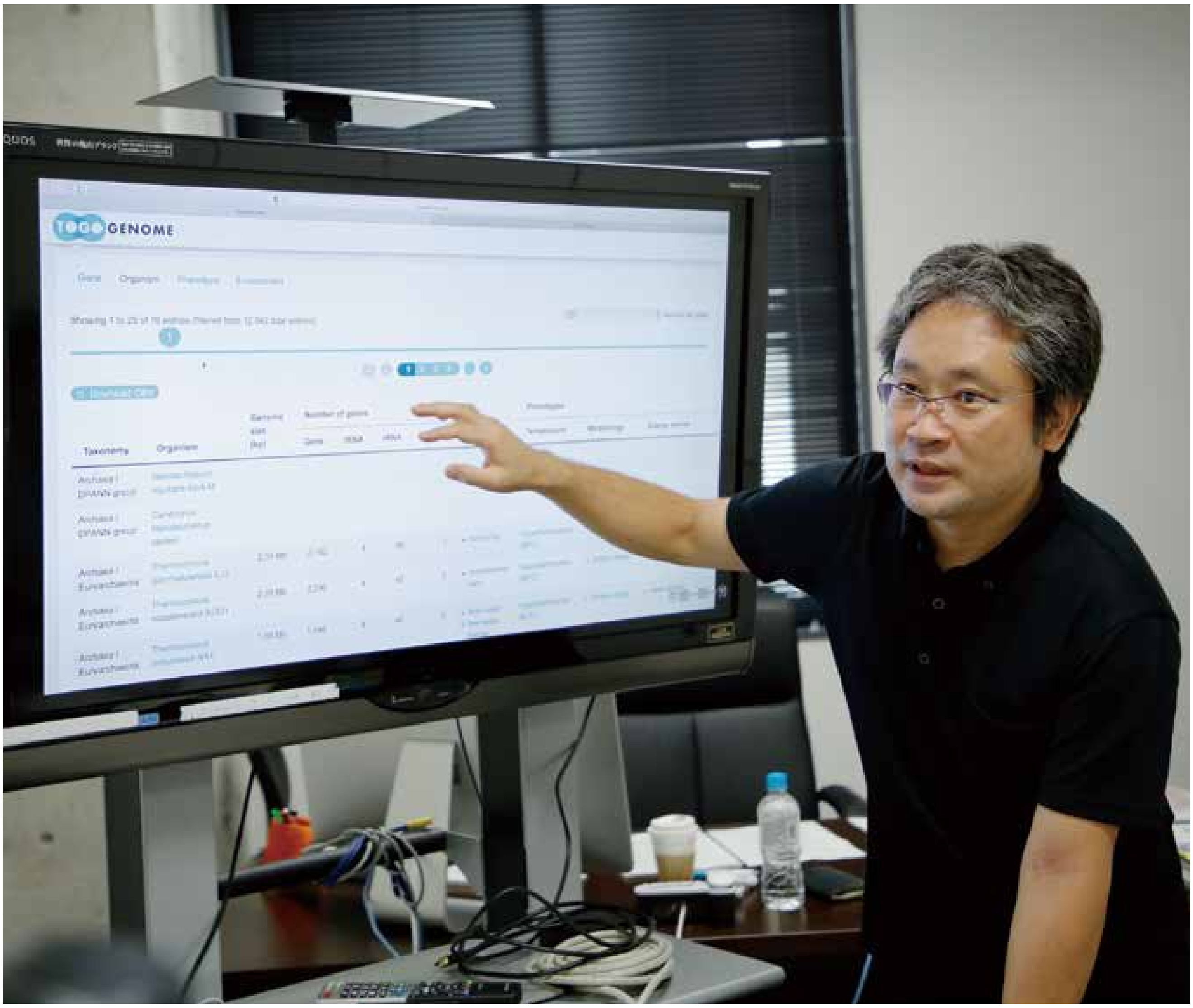
Big Data for Genomic Therapy

For the past 10 years, the DBCLS has organized BioHackathon, an annual international coding camp, to promote standardization and integration of life science data.

"The organizers are particularly focused on the integration of medical databases in recent years because more people are growing interested in connections between genomic mutation and the causes of human diseases and that sort of thing," said Project Assistant Professor Toshiaki Katayama, who has led the hackathon since its inception.

For BioHackathon 2017, Project Assistant Professor Katayama said he expects lively discussions on the topic of genome graphs, a cutting-edge tool for analyzing individual genomes.

"There are slight individual differences in human genomes. Research institutions have traditionally stored variant data separately as reference data. Genome graphs bundle an array of variant datasets into an intelligible batch of information" Project Assistant Professor Katayama said. "This would allow researchers to laser-focus on the genomes of the Japanese, for example. Genome graphs are very useful that way."



Here in the photo, Project Assistant Professor Katayama explains TogoGenome, which is one of the services offered by the DBCLS. He also described about the BioHackathon, which usually consists of a one-day symposium and a five-days hackathon. The symposium was expanded into a two-day event for BioHackathon 2017 to mark the event's 10-year anniversary. "What was new last year was that we had more participants from medical and pharmaceutical fields, including researchers from a Japan Agency for Medical Research and Development (AMED) project. We've devoted the past 10 years for the development of basic technology for database integration. We expect our focus will now shift to international collaborations for standardization in response to the "Global Alliance for Genomics and Health" and other similar initiatives calling for global data sharing," Project Assistant Professor Katayama said. The Symposium portion of BioHackathon 2017 took place during Sept. 9-10 at JST Science Plaza in Chiyoda-ku, Tokyo, followed by the hackathon camp during Sept. 11-16 at Hotel Taikan in Iwate prefecture.

What's also key to clinical use of big data is the development of the standardized "pipelines" that automates a string of tasks from integrating data to connecting genes with diseases to identifying the best medical treatment, Project Assistant Professor Katayama said.

"The pipeline that we envision would use a single graph to bundle all genomic information, allowing the users to apply the same pipeline with the cutting-edge genome graph technology to the subgraph sorted by disease or race," Project Assistant Professor Katayama said.

Project Assistant Professor Katayama pointed out that the development of Common Workflow Language (CWL) used to write such data processing should also become a theme for BioHackathon 2017.

"The entire world is working together to move toward data standardization," he said.

Released on: January 10, 2019 (The Japanese version released on Sept. 11, 2017)

Mapping Out Plant Genomes for Human Survival

Ask an Expert: Satoshi Tabata
Director of Kazusa DNA Research Institute

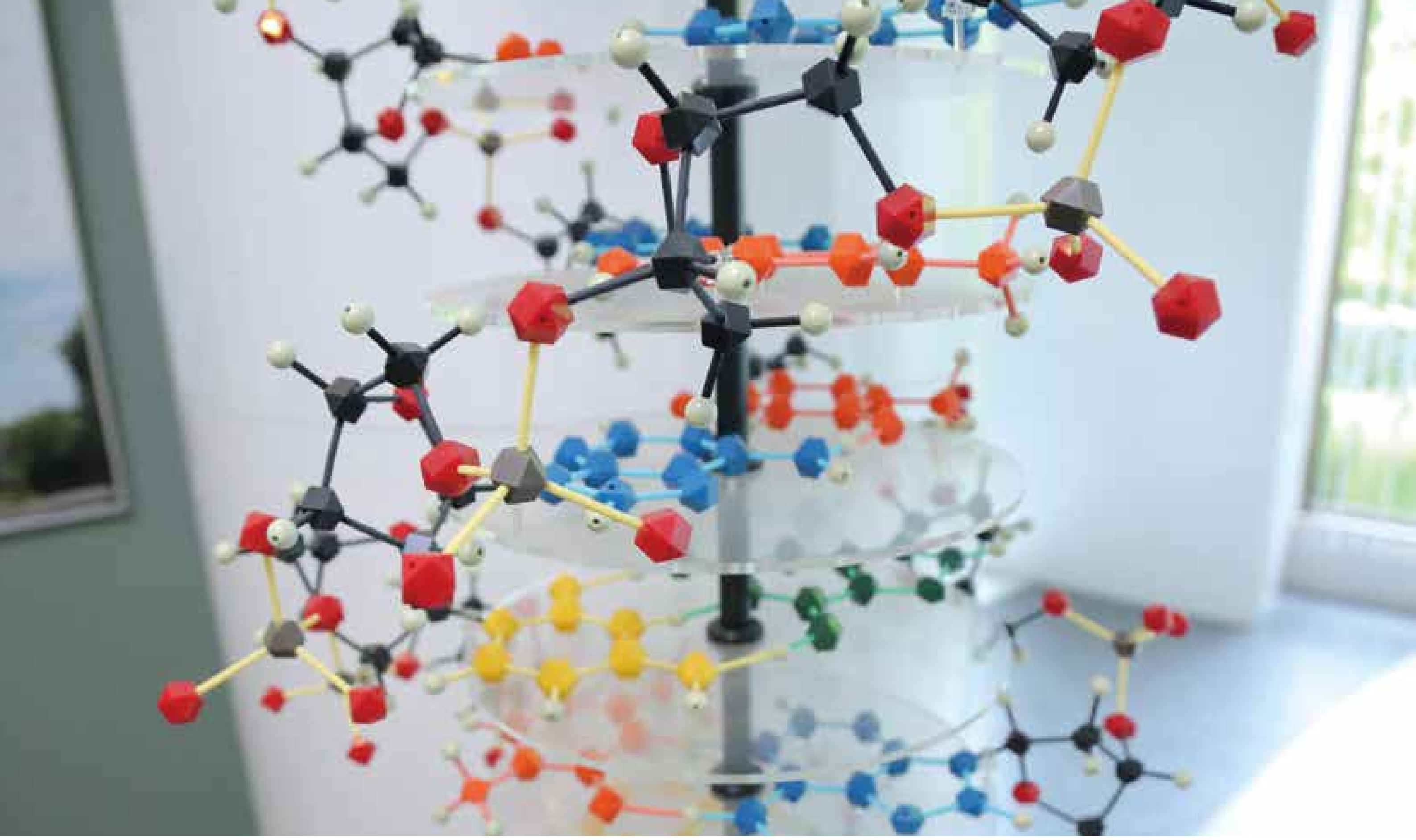
Dr. Tabata is an award-winning genome scientist known for his leading role at Kazusa DNA Research Institute in its successful sequencing of cyanobacteria's complete genome in 1996. He received his bachelor's of science in biology from Kobe University in 1977 and his doctorate of science from Kyoto University in 1983 before serving as a postdoctoral fellow at the University of California, San Diego in 1982 and as a research assistant at Kyoto University Institute of Chemical Research in 1982. He joined Kazusa DNA Research Institute in 1994 as a researcher, where he has served as director since 2013. He has received the gold medal at Tokyo Techno Forum 1997, a Japanese Society of Plant Physiologists Award and a Kumho Science International Award, both in 2001.

If you are trying to make sure everyone on earth will have plenty of quality food to eat, you may care to learn how cyanobacteria's DNA works. The bacteria commonly known as blue-green algae are the origin of chloroplast, and its genomic information provides the fundamental understanding of all plants that depend on photosynthesis to survive. Genomic analysis of such earliest species as well as more recent ones also sheds light on how plants have evolved to overcome ecological challenges over time. Plant genomics is now increasingly gaining the public's attention as weather patterns grow more unpredictable and extreme, necessitating us to tap into the power of life science to find ways to help plants better adapt to the environment. In 1996, Kazusa DNA Research Institute of Japan successfully sequenced cyanobacteria–which was only the fourth completely sequenced genome in the world and the first in Japan. The Institute has since sequenced the genomes of various crop plants, including strawberries, tomatoes, sweet potatoes and napa cabbage, by collaborating with universities, research institutions and corporations. Dr. Satoshi Tabata, the director of Kazusa DNA Research Institute, who played the central role in the 1996 sequencing of cyanobacteria, will explain how researchers could use genomic data of plants for societal common good.

The "Mutants" on Our Dining Table

Plants cannot move like animals do. That explains why plants have more genes and longer genomes than animals do, according to Dr. Tabata of Kazusa DNA Research Institute.

"Plants have numerous genes that can come in to play in order to help adjust to harsh environments, compensating for



their inability to relocate," Dr. Tabata said. "Animals, on the other hand, have fewer genes, which become reprogrammed in response to environmental changes. So, plants and animals have contrasting survival strategies."

Parts of the nucleotide sequences in the genome function as genes while the rest are only there to connect the genes without the protein-producing functionality. Plant genomes contain more gene parts and less connection parts than animal genomes, according to Dr. Tabata. Plants' genomes are also longer.

Many plants also have multiple genomes.

"The human only has one genome comprised of a pair of chromosomes with one side of the pair representing the maternal heredity and the other side paternal. In contrast, some of wheat have sextets of chromosomes, or three genomes. The wheat used to make the bread we eat has three genomes–genomes A, B and D–that are similar to but slightly different from each other," Dr. Tabata said.

But, why does the genome multiply in plants?

"It probably stems from some sort of mutation that occurs when cells divide. Genome multiplication would kill animals, but plants tend to tolerate it," Dr. Tabata said.

"Heterosis–crossbred species' superiority in the growth rate and other features–is also prominent among plants. And we know that genome multiplication causes plants to grow larger. As humans intentionally selected these "mutants" for cultivation, they have become crop plants," Dr. Tabata said.

These crop plants include sweet potatoes that have a sextet of chromosomes and strawberries that have an octet of chromosomes. Their genomes are longer and more complex than animals' genomes.

Kazusa DNA Research Institute's campus in the scenic pasture of the Academia Park in Kisarazu City in Chiba Prefecture.

Dr. Tabata talks about the ongoing research projects at Kazusa, which, in addition to plant genomic research, conducts genome analysis on rare human diseases for personalized medicine per request. The red stuffed animal behind him is CHIBA+KUN, the official mascot of Chiba Prefecture.

"The genomes of crop plants, in particular, are extremely difficult to analyze for that reason," Dr. Tabata said.

From the Lab to the Field

Applying the knowledge gained through genomic analysis inside the lab to crop cultivation is a challenging task.

"In research projects, we've traditionally used 'model plants' that are design-bred for lab uses only. Researchers didn't know how to take the data from lab experiments conducted in meticulously controlled environments and apply that to the field environment where it would suddenly rain and snow and have diseases striking plants out of the blue," Dr. Tabata said.

Dr. Tabata pointed out that this gave rise to the question: "Whose job is it to research breeding techniques for crop plants?"

"I thought we must take on the challenge and establish the linkage between genomics and the data from the fields. To do that, we need to be able to scientifically understand field data and develop methodologies for it. We would then record how plants changed under what conditions and evaluate the data. That's exactly how we are conducting research on crop plants at Kazusa DNA Research Institute," Dr. Tabata said.

The National BioResource Project (NBRP) contains searchable data on about 250,000 different plant species

Arabidopsis thaliana, rice, wheat, barley, chrysanthemum, morning glory, Lotus japonicus, Glycime max / soja, tomato and algae are among the 30 different bioresources that are registered in the National BioResource Project (NBRP) as strategically important for Japan's research projects. The NBRP collects, preserves and shares samples of these bioresources and enables searches of data on 250,000 different plant species on its website. Shown here is a screenshot of the rice database.

"Oryzabase" (rice database)–Core operating organization: The National Institute of Genetics
https://shigen.nig.ac.jp/rice/oryzabase/

The Japan Science and Technology Agency's CREST (Create Revolutionary Technological Seeds for Science and Technology Innovation) is currently promoting this initiative of connecting genomics with field data through its program titled "Creation of fundamental technologies contribute to the elucidation and application for the robustness in plants against environmental changes."

Dr. Tabata believes data and knowledge from plant genomic research should be leveraged to help solve real-world problems, just as human genome research is helping the development of effective treatments for diseases.

"We need to remove some hurdles that are in the way of getting there. I think it's Kazusa's job–and the Japanese science community's job–to figure out how to do just that," Dr. Tabata said. "And this should be a rather immediate goal and not a distant one. Climate is rapidly changing, and we are seeing different pests and plant diseases migrating from Southeast Asia to as far north as the Kyushu Island. We know we need to quickly develop new varieties that can withstand the increasingly challenging environment.

Plant Genomics for Common Good

Dr. Tabata said it's a testament to the advancement of genomic science that people are discussing how to apply the knowledge to cultivation.

"As an undergraduate biology student, I used to wonder how much longer it would take to get the entire genomic picture of a species," he said. "Back then, biology meant analyzing individual phenomena or individual genes that we knew of and compiling those understandings to create a big picture. Then genomics emerged, and what came along with it was this new approach to biology, in which we first list all substances that are useful to understanding the genomic innerworkings of a species and gather data according to the list to develop the whole picture. That's what we call data-driven science," Dr. Tabata said.

For example, researchers would gather and analyze genome-wide data on how different proteins interact with each other, what substances are produced through the metabolic process and what types of chemical chain-reactions those substances have. Kazusa DNA Research Institute is working to expand the databases that allow for such "omics analysis."

"It's important, Dr. Tabata says, given that genomic databases are generally intended for multidisciplinary uses. What researchers need to remember, however, is that multiple factors play into omics analysis; two identical genetic codes in omics data can mean two different things. This is important, Dr. Tabata says, given that genetic databases generally serve as the platforms for multidisciplinary collaborations. Even if you integrate all the data and find a common gene among different kinds of species, it may function differently in each kind of the species because what caused the gene to be born, such as the environmental factors and other genes, vastly vary in each case," Dr. Tabata said. "Genomic data provides the foundation for biological research. How to interpret the data and how to layer that knowledge on top of the foundation will be the key in further advancing plant genomics."

Dr. Tabata has made efforts to advance plant genomics to that end, collaborating for a KAKENHI project called "Platform for Advanced Genome Science (PAGS)" from 2010 to 2014. Using plant genomics for societal common good is important, Dr. Tabata said. After all, the survival of plants means the survival of humans.

Released on: May 31, 2018 (The Japanese version released on Sept. 11, 2017)

What Do Chimps' Mutant Genes Say About Us?

Ask the Experts: Tetsuro Matsuzawa
Distinguished Professor of Kyoto University Institute for Advanced Study (KUIAS)

Dr. Tetsuro Matsuzawa serves as deputy director-general and distinguished professor at Kyoto University Institute for Advanced Study (KUIAS) with a joint appointment as professor for the university's Primate Research Institute. In November 1977, three and a half years after receiving his bachelor's degree in philosophy from Kyoto University in 1974, Dr. Matsuzawa founded a research project known as Ai Project to study the mind of the chimpanzee while also conducting surveys on the life of chimpanzees in the wild. He has used his research result to better understand human intelligence and behaviors as well as evolution, establishing a new academic field called comparative cognitive science.

Ask the Experts: Asao Fujiyama
Specially-appointed Professor of National Institute of Genetics (NIG)

Dr. Asao Fujiyama serves as the director of the Joint Support-Center for Data Science Research (DS) at the Research Organization of Information and Systems (ROIS) and as a specially-appointed professor for the National Institute of Genetics (NIG). He holds a Ph.D. in science from Nagoya University. He worked at the Faculty of Medicine at Osaka University, the Institute for Molecular and Cellular Biology at Osaka University, Cold Spring Harbor Laboratory and the University of Chicago before becoming assistant professor at the NIG in 1989, professor at the National Institute of Informatics in 2002 and professor at the NIG in 2008. He joined the DS as its director in 2016. He specializes in genomic bioscience and is known for comparative analysis of human and chimpanzee genomes.

© Primate Research Institute, Kyoto University

Ayumu is an 18-year-old chimpanzee living in Japan who, much like his mother, works hard on any task given to him until he gets it done. He also has handsome looks that came from his father. But how do these traits relate to his genetic makeup? Which genes did Ayumu inherit from his father, and which ones from his mother? And how likely is it that Ayumu has mutate genes that weren't present in his ancestors or even in his grandparents?

A group of Japanese scientists now have answers to all the questions. In a press conference at Kyoto University on Oct. 30, 2017, Dr. Tetsuro Matsuzawa from the Kyoto University Institute for Advanced Study(KUIAS), Dr. Asao Fujiyama from the National Institute of Genetics (NIG) and Dr. Yasuhiro Go, associate professor for the Center for Novel Science Initiatives at the National Institutes of Natural Sciences (NINS), announced that their research team had completed an "ultra-deep" whole genome sequencing study on Ayumu and his parents. The chimpanzee trio live at the Primate Research Institute of Kyoto University in Inuyama City in Aichi prefecture. The ultra-deep study, which analyzed an unprecedented number of DNA bases in the primate's study history, revealed each offspring of the chimpanzee has estimated 1.48 newly-formed mutations per 100 million bases of DNA.

Dr. Matsuzawa and Dr. Fujiyama will explain what the results of the study on the chimpanzee family at Kyoto University's Primate Research Institute say about chimpanzees' closest cousin, the human, and the future of genome studies.

Genomes, the Driver of Behaviors

The Primate Research Institute of Kyoto University, a worldly renowned primate research organization, is home to an extended chimpanzee family spanning three generations. Prof. Matsuzawa has been studying "the mind of the chimpanzee" since 1977 when he met Ai, a girl chimpanzee who had moved to the center earlier that year. Ai was estimated to be 23 years old when she gave birth to Ayumu on April 24, 2000. In his book, "Okāsan-ni natta Ai" ("Ai who became a mom"), Prof. Matsuzawa described the details of the labor, including how Ayumu's dangling limbs had researchers concerned about his health at first.

Prof. Tetsuro Matsuzawa (right) and Prof. Asao Fujiyama during an event at Kyoto University

"We have been conducting research on the learning behaviors of chimpanzees to find out how knowledge and experience accumulate to make their minds grow and change," Prof. Matsuzawa said. "If I can use the iceberg metaphor, the tip is the behaviors, and the foundation that we cannot see under the sea is the genomes. Once we understand exactly how genes work, then the whole picture of the iceberg will begin to emerge more clearly," he said. "For example, we already know that aggressive behaviors in chimpanzees have to do with a male hormone called Androgen and which genes function as Androgen receptors. This leads to our next question about which genes can help suppress aggressiveness and so on, and we can continue digging down into the matter this way."

Going Back 6 Million Years to Fill the Genetic Gaps

Humans and chimpanzees share so much of their DNA. This made Prof. Fujiyama decide to delve into the chimpanzee's genome in the early 2000s while participating in a human genome project.

"By comparing the human genome to that of the chimpanzee that's phylogenetically closest to us, we can figure out the genes that played the crucial roles in separating the two species from the common ancestor. And because most of the viruses that cause infections for chimpanzees also affect humans and vice versa, researchers are hopeful that studying chimpanzee genomes will help them develop effective treat-

ments for human diseases," Prof. Fujiyama said.

Prof. Fujiyama and fellow researchers have already conducted some ground-breaking research in the field, identifying mirroring relationships between the chimpanzee's 22 chromosomes and the human's 21 chromosomes and studying their commonalities and differences.

"Following our study in 2004, we concluded the human and the chimpanzee phylogenetically parted ways about 6 million years ago. We simulated the evolution on the premise that the generation time is 20 years and computed the rate of genetic mutation," Prof. Fujiyama said.

Prof. Matsuzawa won't forget the day Prof. Fujiyama visited him in Inuyama City to collaborate on that study in 2004.

"Ninety-nine out of 100 people who visit the Primate Research Institute look at the chimpanzees here and ask which one Ai is. Only Prof. Fujiyama asked, 'Which one is Gon?'" Prof. Matsuzawa said, laughing. Gon is a root node of the Institute's chimpanzee pedigree tree. "Prof. Fujiyama's question showed that back in 2004, he already knew how to go about the chimpanzee genome study," Prof. Matsuzawa said.

And, that's how the ultra-deep whole genome sequencing of Ai, Ayumu and his father, Akira, began.

Prof. Matsuzawa believes the lack of research on the mind of the horse has to do with 'Clever Hans,' a horse that became famous for its claimed ability to perform math around the turn of the 20th Century," Prof. Matsuzawa said. "A researcher eventually proved that it was all tricks. Little study has since been conducted to understand how the mind of the horse works, even though people have continuously used the ability of horses to take directions for horse races and other purposes."

Taking the Guesswork out of Genomic Analysis

The whole genome sequencing on Ayumu and his parents was different from any genome studies conducted in the past on two fronts. First, the team ran a "personal genomics" study on the chimpanzee nuclear family, treating the parents-offspring trio as a set. Secondly, the team used the so-called next-generation sequencer to analyze 575, 463, and 468 giga bases of DNA–unprecedented numbers of reads for personal genomic sequencing of any mammalian species–in order to make the analysis highly accurate. Each chimpanzee has about 30 giga bases of DNA. The study also considered the actual generation time instead of an estimate used in the traditional phylogenetic research.

"From this research, we were able to determine the rate of germline newly formed mutations to be 1.48 mutations per 100 million bases," said Prof. Go. "This is higher than the reported rate of 0.96 to 1.2 mutations per 100 million bases for humans. We also discovered 75 percent of the newly formed mutations for the chimpanzee originated with the paternal side (sperm)."

In other words, the team was able to pinpoint the types, the extent and frequency of errors in the transmission of genetic information from parents to an offspring.

"If we can go beyond the parents-and-son trio to cover more members of the second generation or even the third generation in the analysis, we will have a greater understanding of the inner workings of the chimpanzee's genome," Prof. Fujiyama said.

Tracing the Mammals' Footsteps

For Prof. Matsuzawa, the next step in his research is to target non-primate mammals as study subjects. He is particularly interested in understanding the mind of the horse in connection with the genome at the bottom of the "iceberg." He plans to conduct research both in laboratory settings and in the wild just as he's done for his chimpanzee studies.

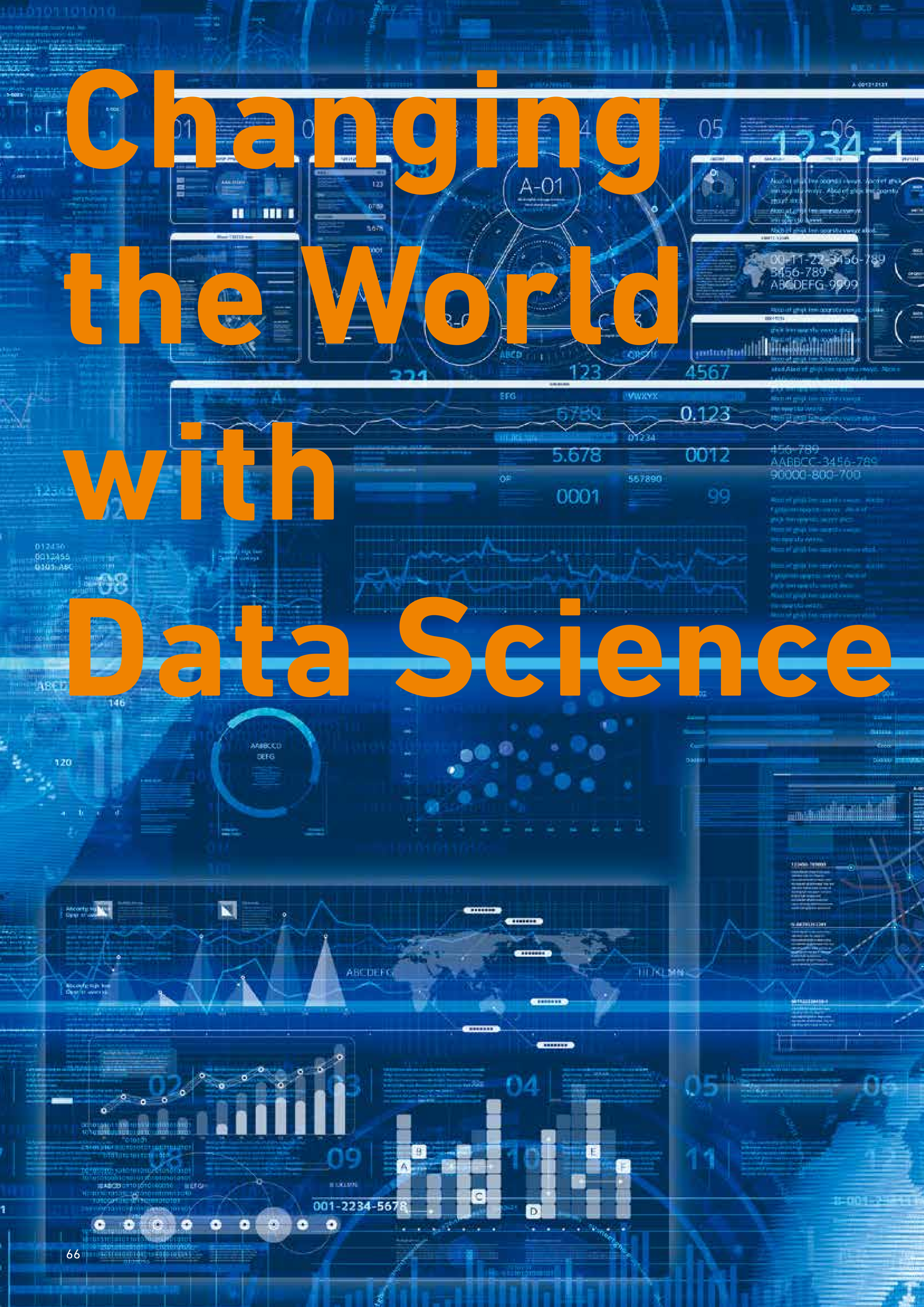
"The horse is a beloved animal, yet no one is studying the mind of the horse. It's like a blind spot in our field," Prof. Matsuzawa said.

He pointed out that studying mammals helps researchers better understand the primate and that the genome is the key to the big picture.

"Following the extinction of dinosaurs 66 million years ago, mammals adapted to the new environment and rapidly proliferated. Mammals are believed to have all come from a small, nocturnal creature resembling a mouse. This explains why many mammals are terrestrial," Prof. Matsuzawa said. "In the meantime, the primate chose to live in the tree. As they began climbing trees, their four legs became four hands. Take a look at the four limbs of chimpanzees or Japanese macaques or whatever, and you will see they are hand-shaped. But the human is unique in that they decided to return to the ground from trees. We got our feet when we became humans," he said. "When you think these things, you start to see the history of evolution. And, the genome is behind all of it."

Released on: July 10, 2018 (The Japanese version released on Sept. 11, 2017)

Changing the World with Data Science



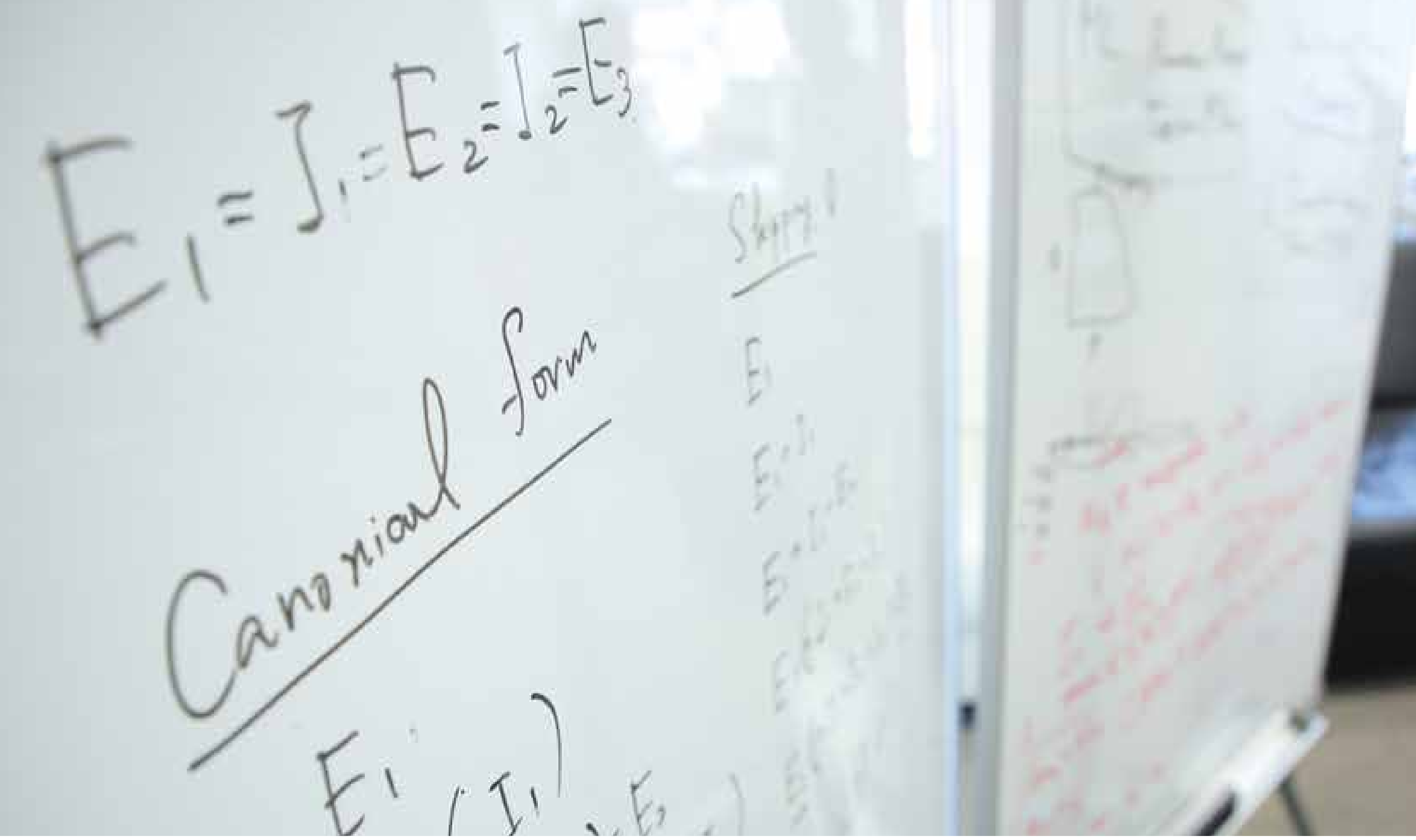
Reinventing the Process of Making Things

Ask the Experts: Ryo Yoshida
Professor of The Institute of
Statistical Mathematics (ISM)

Dr. Ryo Yoshida, professor for Department of Data Science at the Institute of Statistical Mathematics (ISM), has served as the director of the Institute's Data Science Center for Creative Design and Manufacturing since the center's opening in July 2017. After receiving his doctorate degree in Statistical Mathematics from the Graduate University for Advanced Studies in 2004, he worked as a Project Assistant Professor for the Human Genome Center at Institute of Medical Science, the University of Tokyo–a position he has maintained after joining the ISM in 2007. In addition, he serves as a specially appointed researcher for National Institute for Materials Science. Dr. Yoshida has the experience of using his expertise in data science for research work in both biology and materials science. He plays leadership roles in several cutting-edge research projects, including the Japan Science and Technology Agency (JST)'s the "Material Research by Information Integration" initiative.

Ask the Experts: Ichiro Hasuo
Associate Professor of
National Institute of Informatics (NII)

Dr. Ichiro Hasuo, associate professor for the National Institute of Informatics (NII)'s Information Systems Architecture Science Research Division, has served as the director of the Institute's new Global Research Center for System Design and Mathematics since July 2017. With his expertise in theoretical computer science, he is particularly interested in such research topics as system verification, programming language theory, information processing, and mathematical structures in information science. He is a 2002 graduate of the University of Tokyo and received his doctorate in computer science from Radboud University Nijmegen in the Netherlands. Before joining the NII in April 2017, he served as associate professor for the Graduate School of Information Science and Technology at the University of Tokyo as well as visiting associate professor for the Research Institute for Mathematical Science at Kyoto University.



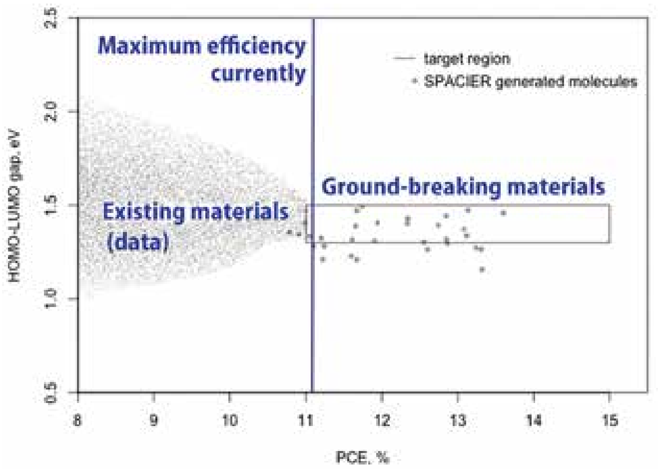
Decades after “Made in Japan” became synonymous with high quality, the country is pushing for scientific breakthroughs that would help industries redefine excellence in manufacturing. The initiative focuses on leveraging data science and artificial intelligence (AI) for manufacturing, following in the footsteps of the U.S., Germany and many other nations in Europe and Asia, where similar movements have already been underway. In hopes of galvanizing ongoing efforts in Japan, the Research Organization of Information and Systems (ROIS) established two centers last year within the existing institutes under the ROIS’s umbrella–the Data Science Center for Creative Design and Manufacturing, launched at the Institute of Statistical Mathematics, and the Global Research Center for System Design and Mathematics, launched at the National Institute of Informatics. So, what’s the vision of Japanese leaders in this scientific frontier? And how could they take manufacturing to a whole new height? The scholars who lead the two new research centers will spell it out for us as they explain about the projects that they lead.

Materials Informatics: Trading Test Tubes for AI

One of the study fields that are increasingly attracting the attention of both the academic and manufacturing communities is materials informatics. Public interest in the field grew following the 2011 launching of the Materials Genome Informatics Initiative by the U.S. government under the leadership of the former president Obama.

Researchers used to go through countless trials and errors in wet labs to develop new materials for products. When things do or don’t work out in the ways hypothesized, that would shed that much more lights on the complex relationships between material chemistry and physics. Researchers would use the knowledge to conduct additional experiments and inch toward discovering ground-breaking materials. This long-standing empirical approach to inventing materials is becoming a thing of the past, now that materials informatics allows researchers to make discoveries without compiling direct evidence to get there, according to Prof. Ryo Yoshida of the Institute of Statistical Mathematics (ISM).

Materials informatics refers to a field of study that har-



The graph shows newly discovered materials for photovoltaic cells. The square inside the graph, in which few dots exist, indicates the area where researchers are targeting.

Prof. Yoshida is also interested in designing of spaces conducive to industry-academia collaborations. While working with companies on projects, Prof. Yoshida helps them build mechanisms for recruiting talented data scientists.

nesses the power of big data and AI to rapidly discover novel materials that could be created. Before conducting direct experiments and physically measuring values, researchers would put existing big data as well as interpolative data derived from it through algorithms and let AI identify patterns in the data. AI uses the patterns to predict how proposed materials would behave in the real world and/or what properties a material need to be made of to offer desired characteristics and features.

This drastically accelerates the speed of material innovations. It also allows researchers to create something that they never imagined or have considered as far-fetched, said Prof. Yoshida, who serves as the director of the ISM's Data Science Center for Creative Design and Manufacturing.

For example, the Power conversion efficiency of photovoltaic cells is currently limited to 11 percent maximum, Prof. Yoshida said, as he showed a graph with dots representing the efficiency of cells made of different materials.

"The vertical line in the middle is the maximum efficiency we can attain with the existing scientific knowledge and materials. The area to the right is the 'uncharted territory.' Our job is to find materials that fall in this territory," Prof. Yoshida said. "We want materials with features and characteristics that enable cells to achieve far more than 11 percent of efficiency," he said.

"Over many decades, material science gradually advanced, pushing the middle line in the graph to the right. Now, we have the opportunity to expand this dotted region at exponential speeds by taking advantage of the cutting-edge data science technology. This will definitely be a boon to scientific and industrial advancement."

Prof. Yoshida pointed out that data-gathering and identification of patterns with the use of machines is the essence of data science.

"In data science, we, the researchers don't have to come up with any theories. We just need to have the machines analyze data to recognize hidden patterns and identify the theory behind them with inverse problem methods. The computer uses the theory to build a virtual material with the structure and properties that are necessary to provide desired features," Prof. Yoshida said.

Getting Mileage out of Data Science with Extrapolative Data

Prof. Yoshida has particularly been interested in the use of extrapolated data to enhance the machine's ability to create a virtual material far, far outside of the current realm of possibility. That's because the computer needs more data than currently available to theorize advanced materials on it.

"You can make only so much prediction from data. To explore outside the 'predictable zone,' you would need additional data. That's the limitation of data science," Prof. Yoshida said.

"Data science has never played crucial roles in any Novel Prize-level, ground-breaking scientific discoveries. Technological advancements have always come from direct experimentations and theories. When I think about what contributed to my research successes thus far, data science accounts for 5 to 10 percent of it," Prof. Yoshida said.

He noted that data science has increasingly had artistic applications in recent years, such as machine production of videography and music and designing of anime characters. However, whatever the field you are in, the availability of data or the lack thereof still dictates what you can do with data science," Prof. Yoshida said. "Therefore, if we wanted a machine that can make advanced predictions–those that go far beyond the predictions based on the knowledge and experiences humans currently have, then the machine needs to be capable of extrapolating data to make up for what's missing."

This idea inspired Prof. Yoshida's research team to develop an algorithm called SPACIER.

"SPACIER sets up a virtual materials research laboratory inside a computer. It designs experiments and conducts simulations in the virtual lab to extrapolate data," Prof. Yoshida said. "The machine solves inverse problem to enhance its predictability, so it can find materials beyond the existing 'predictable zone.' You can repeat this extrapolation process through the algorithm to gradually push the limit."

Theoretically, he said, it's possible to keep using the algorithm to generate more and more data, further expanding data distributions. It's also possible to modify SPACIER for general applications in other fields, including such a purpose as structural designing of industrial products, Prof. Yoshida said.

Collaborations and Blending Tools Key to Reaching the Goal

While combinations of algorithm, AI and extrapolative data may sound like a magic wand for inventing things, Prof. Yoshida cautions that there's still a way to go before practical applications become routine. Don't get too excited about machines like AlphaGo Zero, the robot that gets better at playing an Asian board game known as Go by learning from its own actions to outdo itself.

"We've only begun making 'super creative' machines that internally generate data to serve its purpose. There are many technological challenges to overcome to perfect the mechanisms," he said. "At least for now, experiments in virtual labs alone cannot suffice to meet all industrial design and development needs. It's important that we create formulae for turning ideas into products by blending physical experiments, theories, computational research and algorithms."

Prof. Yoshida's mid-range vision calls for robust machine learning with the use of data and theories from direct experiments. A machine would feed on existing data and take back what's come out of the algorithm after theories are applied to it. But, how would he convince the manufacturing industry to adopt the new methodology?

"Show. Don't tell," Prof. Yoshida said. "We want to send the message out that the methodology does work and that data science can be the most powerful ally in innovating products under the right partnerships. I intend to bring this message to those on the front line of industrial development research over the next few years."

As part of this initiative, the Data Science Center for Creative Design and Manufacturing plans to collaborate with six corporations this year for data scientists training. The center hopes to work with many more companies in the years to come, Prof. Yoshida said.

New Era of Quality Assurance Begins with Self-Driving Cars

Another area in information science that a group of ROIS researchers is trailblazing to advance manufacturing is called "category theory." This effort stems from the fact that most machines today are computerized, necessitating manufacturers to reconsider how to carry out quality assurance, according to Prof. Ichiro Hasuo of the National Institute of Informatics (NII)'s Information Systems Architecture Science Research Division.

"Just think about self-driving automobiles," said Prof. Hasuo, who also serves as the director of the NII's Institute's new Global Research Center for System Design and Mathematics. "Even the most experienced auto makers' considerable know-how can't be of enough help when it comes to performing quality control and assurance in this nascent field of self-driving automobile technology. Manufacturers are understandably unsure about how to navigate these uncharted waters. Academia are equally confused about which discipline's responsibility it should be to address this issue."

That's where Japan Science and Technology Agency (JST)'s Exploratory Research for Advanced Technology Metamathematics for Systems Design Project (ERATO MMSD project) comes in. Directed by Prof. Hasuo, the ERATO MMSD project aims to expand the formal methods–mathematical methods to structure software systems–with the use of category theory, so formal methods can be applied to cyber-physical systems (CPS) as well as physical products. The project places a particular focus on automobiles.

"It began at the right timing," Prof. Hasuo said of the ERATO MMSD project, which was launched in October 2016 and is now housed at the Institute's new Global Research Center for System Design and Mathematics, for which he serves as the director.

"Industrial equipment used to be kinetic machines, and their stabilization systems were based on the control theory. But nowadays, nearly every machine has a computerized control system installed in it," Prof. Hasuo said. "Generally,

simpler machines are linear systems, meaning their response to outside forces is straightforward and predictable. If you give it twice as much force, the output would also be twice as much. Computerized control systems for automobiles and other large, complex systems don't act that way."

So, how do they act? The ERATO MMSD Project will answer the question, Prof. Hasuo said.

Category Theory for New QA Methods

The ERATO MMSD project couples logic with category theory to develop a method that manufacturers can use to assure the safety of their products.

"Category theory comes from algebra. It's a mathematical language used for structural description," said Prof. Hasuo, a category theory specialist. "It focuses on the relations among objects and expresses it in an abstract form. It differs from 'graph theory,' which also describes the relations among objects, in that category theory doesn't generalize what's present; it abstracts the logics behind what's present. In other words, category theory is a meta language."

Category theory is an integral part of mathematical structuralism. The theory matured at the same time as structural anthropology grew out of the study of algebraic structures in France during the 1940s. This historical fact may make one wonder if and how category theory and structural anthropology–which owes its development to Claude Lévi-Strauss–may be related to one another.

"Claude Lévi-Strauss (French anthropologist whose work helped the development of structuralism) made many contributions to the academic world, but I understand one of them was his idea that all kinds of communities–or mass bodies of things–have immutable structures supporting them. I believe the same is true for masses of data. Whether it's a traditional information system–which is the focus of the data science–or an information system involving physics data of industrial machines, they both deal with collections of information. So, they've got to share some things in common from an algebraic perspective," Prof. Hasuo said.

JST ERATO Metamathematics for Systems Design Project is based out of a downtown Tokyo building. A group responsible for expansion of formal methods for general usage and a group responsible for applying the methods for practical purposes converge here daily and have lively discussions.

Mathematic Processes as Guide for Exploring Unfamiliar Territory

Prof. Hasuo's is not the first researcher to suggest the use of formal methods as verification techniques for physical information systems. In fact, there have been many attempts made around the world to do just that. "Those attempts have not necessarily been unsuccessful because it is not clear how to make formal methods usable for cyber-physical information systems," Prof. Hasuo said. He wants to use category theory to create the missing link, and that makes his idea and project exceptionally cutting-edge.

Formal methods–which are used to verify software does its job in the way it's supposed to–are regarded as reliable quality-assurance techniques because they are mathematical, and therefore, leave little room for ambiguity and error. For manufacturers, minor inaccuracy could mean a disaster. Hair-splitting margins of error have indeed caused problems, such as rocket launching problems and arithmetic glitches in computers. Formal methods are useful for ensuring software products won't get stuck in an infinite loop nor present any vulnerability. So, it could change the ways industries operate around the world if the ERATO MMSD project succeeds.

"Mathematical processes show you the ropes to guide you through the unknown territory," Prof. Hasuo said. "We are going to use category theory to expand formal methods and create templates for defining problems in a mathematic language. We will gradually flesh them out and keep the templates in our 'tool box,' if you will, so we can take out suitable one, depending on what problem we need to solve, and work with manufacturing engineers to provide evidence for whatever they need to verify or prove. The ever-advancing AI and machine-learning technology can be useful for this process, as well," he said. "When this is put to use, this will provide additional support for the current quality-assurance process centering around engineers' judgment calls based on their years of experience and knowledge. This should drastically reduce the time required for verification processes."

Prof. Hasuo and his team want to see five cases of manufacturing applications of this verification method before the ERATO MMSD project completes in 5 ½ years. The team is already discussing possible applications with several companies, according to Prof. Hasuo.

Released on: August 10, 2018
(The Japanese version released on January 10, 2018)

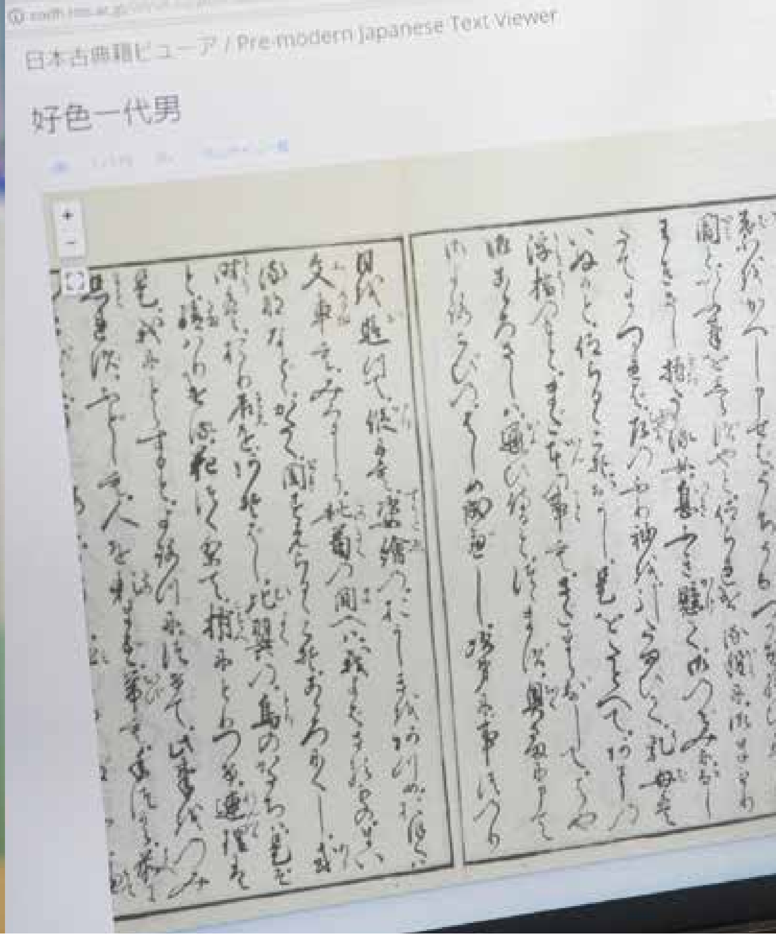
Data Science–the Humanities' New Tool

Ask the Experts: Naruya Saitou
Professor of
National Institute of Genetics (NIG)

Prof. Saitou serves as professor of genetics for the Graduate School of Advanced Studies and as professor of biology at the University of Tokyo. He has extensive experience in research on species-specific evolutionarily conserved characteristics based on large-scale genomic data comparisons. His 1987 paper with Dr. Masatoshi Nei, his Ph.D. supervisor, which proposed the "neighbor-joining method" for reconstructing phylogenetic trees through the use of evolutionary distance data, has thus far been cited more than 53,000 times. He is the author of "Introduction to Evolutionary Genomics, Second Edition (2018, Springer)", and many other books.

Ask the Experts: Asanobu Kitamoto
Director of Center for Open Data in the Humanities (CODH), Research Organization of Information and Systems (ROIS)

Prof. Kitamoto, the director of the Center for Open Data in the Humanities at the Research Organization of Information and Systems, serves as associate professor for the Digital Content and Media Sciences research division of the National Institute of Informatics and as associate professor of informatics at the Graduate School of Advanced Studies. He has a bachelor's and doctoral degrees in engineering from the University of Tokyo. His research work centers around image data analysis and application of data-driven science in various disciplines from the humanities to earth science to disaster reduction. His project was among Jury's Selections at the 2007 Japan Media Art Festival organized by the Agency for Cultural Affairs and the Information Processing Society of Japan's Yamashita SIG Research Award. Prof. Kitamoto is interested in transdisciplinary research that helps promote open science.



History teaches us something new whenever we revisit it, but researchers are finding out that we can learn so much more from our past by bringing data science into historical studies. From projects that combine genomics and archeology to digitization of cultural heritage, a wide range of experiments that merge the humanities and computer science are bearing fruit, shedding light on previously unknown facts. This is allowing researchers to dig deeper into their subject matter to shed light on previously hidden facts–so much so that historians, for example, are now able to add new pages to history books or even rewrite parts of them. Two data science experts involved in this growing research approach explain how this is helping us develop new bodies of knowledge and where it's all headed.

Upside of Drilling into Ancient Artifacts

When and how did the ancestors of the modern Japanese arrive the archipelago from the Asian continent?

It's a question that fascinates the people of this island country in Far East. A piece of enlightenment came in 2016 when Prof. Naruya Saitou of the National Institute of Genetics (NIG) and his research team analyzed the genome of human bones that had been excavated from the Sangaji shell mound in Fukushima, Japan, and found that the Jomon people–who lived in the Japanese Archipelago between 14,000 and 900 BC–were genetically very different from Chinese and Southeast Asians. The team made the discovery upon visiting the University of Tokyo Museum where the bones were stored and drilling a hole into a back tooth of both a male and female skull to take samples.

Then in 2017, the team concluded that their Japanese ancestors from the continent arrived in three separate time frames instead of two, which had been the popular belief.

"As we compared the genome sequences of people who live in various locations of the Japanese Archipelago (Yaponesia), we found small differences among them.When we looked at where these groups were located, we saw a geographical pattern. And that looked different from the pattern of the Jomon population distribution expected under the two-step migration model," Prof. Saitou said. "As you can see, humanities information is critical to advancing the research on human evolution. We have countless pieces of archeological artifacts painstakingly unearthed from across the country over many, many decades. Our job is to mine these materials for data. In other words, museums are our 'field' of research," Prof. Saitou said.

He pointed out that such "genomic excavation" is happening more and more across Japan.

"Genomics is finally catching up to the humanities. As we continue with the large-scale genomic project that studies the Jomon people as well as modern-day Japanese from many different areas of the country, we will be able to gain more insights into the revolutionary history," he said.

Data at the Crossroad of the Humanities and Science

Bringing together the humanities and data science–the two seemingly opposite ends of the academic spectrum–for the advancement of both is an emerging trend. But, Prof. Saitou said a 19th Century scientist's concept of how the universe

works to is the most useful for explaining why the new approach makes perfect sense. What Prof. Saitou is referring is a Venn diagram that Japanese natural historian and folklorist, Kumagusu Minakata, drew in his letter to a priest. One of the two circles in Minakata's diagram contains a character that signifies "the mind," with "matter" encircled in the other. The overlapped area represents "things."

"There is the human mind, or the subjective world, on one side and physical matters and phenomena, or the objective world on the other. Both have 'things' in common, and that's what we obtain through excavation, such as historical data and archives," Prof. Saitou said.

"People talk about how far biology has come, but there is so much more that we don't know," said Prof. Saitou, who works to advance "evolutionary genomics," a field that studies how the genome has changed over the course of evolution.

"An interesting thing about genomics is that the 'matter' and 'things' mirror each other. While adenine (A), cytosine (C), guanine (G), and thymine (T) are genetic information, or 'thing,' each of them also corresponds to a physical, objective 'matter.' This closeness between matter and things is a wonderful thing," Prof. Saitou said.

"Our goal is to find out how a natural phenomenon called evolution transpired over time by first looking at the genomic changes. It's important for us to use data as the language to describe what happened in the world that surrounds the human and how it changed. This enables us to prove or disprove history, and I believe that's what data science is all about," he said.

Why Data Speaks Against Darwinism

In the research field of biological revolution, Prof. Saitou said, tracking changes in genomic data has already allowed scientists to disprove one of the most widely held ideas: the Darwin theory.

Charles Darwin believed that every species' survival is dictated by its ability to adapt to the environment through genetic mutations. In other words, genes are continually replaced by those that can better handle environmental challenges, allowing species to produce more offspring and carry forward the mutations. Prof. Saitou calls this theory of natural selection an "illusion," however.

"The idea that all biological organisms are adjusted to the environment is a fantasy. If better genes outlive worse ones, then no species should die off. But in reality, we see species go extinct all the time. Just this fact itself disproves the natural selection process," Prof. Saitou said.

He said a close look at genomic data further provides evidence that the theory of natural selection isn't factual.

"Even when mutations occur in a species' DNA to better handle the environment, the effects eventually wear off, and those mutated genes will play no more or less part than other genes in the species' survival. In the end, maintaining the status quo is all that the genes do. We call it 'purifying selection' as opposed to 'natural selection,'" Prof. Saitou said.

In 1968, Japanese biologist Motoo Kimura proposed the "neutral theory of molecular evolution," stating that the reproductive rates of organisms remain unchanged, or "neutral," even if nucleotide sequences become altered as the result of mutations. This remains the most trusted theory in the field of evolution studies, Prof. Saitou said.

"Today, we know most genes have gradually evolved from the originals as such neutral mutations repeatedly occurred over long periods of time," Prof. Saitou said.

Data for Reconstructing the Edo Period

On another front of transdisciplinary data science is the academic community's effort to pull together all types of historical archives to build big data and analyze it to connect all the dots. The Center for Open Data in the Humanities (CODH) at the Research Organization of Information and Systems (ROIS) supports this movement by offering datasets, software and other tools to innovate humanities research. In CODH's March 2018 seminar, "Historical Big Data - Challenges in Transforming Historical Documents to Structured Data for the Integrated Analysis of Records in the Past -," for example, participants from various disciplines learned how to integrate and analyze seismologic, climatologic and astronomical data obtained from historical records.

"People didn't have Twitter in old days. Instead of posting on social media, they wrote down in their documents and diaries what they saw and heard. Historical manuscripts and old records are a treasure trove of such everyday information from the past," said CODH Director Asanobu Kitamoto, who also serves as associate professor for the Digital Content and Media Sciences research division of the National Institute of Informatics and as associate professor of informatics at the Graduate School of Advanced Studies.

"For example, we can find weather information in a 300-year-old diary that a family in the western part of Tokyo has kept. In Nagano, people have maintained official records of "Omiwatari" for the past 600 years, which is a religious celebration of the appearance of an ice ridge formed along a crack on a frozen Suwa Lake. Shrines in Kyoto have also long recorded the day cherry flowers blossomed every spring," Prof. Kitamoto said.

He said records from the Edo period are particularly abundant, which is the reason he is focusing his research on that period. Researchers can use such records to reconstruct past climate conditions, put together historical disaster information and even analyze correlations between natural events and social, political or economic events, including market fluctuations in the modern era, and groundwater management. The CODH works to foster the research community of digital humanities by supplying its technological expertise and research gathering.

"Reconstruction of history through comprehensive analysis of historical records has been a growing movement worldwide. The Venice Time Machine, a project to create digital archives of records covering Venice's 1000-year history to reconstruct the city from the ancient days, is an example," Prof. Kitamoto said. "But research teams engaged in these digital humanities projects are scattered across the globe. We are hoping to establish a system for bringing together the research results, digitizing them and making them accessible to all researchers," Prof. Kitamoto said.

Open Science Through Standardization for Visual Art Curation

The CODH is also an active participant in the International Image Interoperability Framework, or IIIF, which is an international initiative to develop standard technology and systems for image delivery. IIIF aims to make digital images–those that belong to museums, libraries and other institutions and individuals from around the world–accessible to researchers and the general public everywhere. The Nation-

The CODH also offers a publicly assessible collection of the images of characters from 15 old books that belong to the National Institute of Japanese Literature. The collection contains 403,242 characters representing 3,999 different cursive characters. "Ruby characters (annotations for pronunciations) are particularly a challenge for AI in recognizing characters," Prof. Kitamoto said.

al Diet Library of Japan, the British Library, the National Library of France and "Europeana," an online repository of digitized items, as well as Oxford University and other universities, use IIIF to provide access to more than 350 million digital images. As part of its effort to help build a IIIF community in Japan, the CODH has developed an application called "IIIF Curation Viewer," which allows users to view IIIF images on a global scale, including the National Institute of Japanese Literature's old books on the CODH's website.

"IIIF Curation Viewer lets you cut out the parts of images that interest you, such as people's faces, and curate them just like making a scrapbook. To do something like this, we used to visit museums, take photos of the images or photocopy them and cut and paste the parts we want with scissors and glue. Now that we have this Viewer, we can do the same thing hundreds or thousands of times faster," Prof. Kitamoto said.

As curation and analysis of images becomes easier, researchers are tapping more into visual resources to make discoveries.

"We curated facial images from our collection as an experiment. Comparison of these images revealed that a face appears in one book is painted very similarly to another face that appears in a different book. Picture books from the Edo period are comprised of paintings and calligraphy, and we know people used to ask art studios that specialize in either painting or calligraphy to get each part done separately. It is possible that there were templates that art studios were using to paint faces–which would be somewhat similar to the method that today's manga artists use," Prof. Kitamoto said.

He noted that everyone in the research community significantly benefits from open access to research data, as it allows them to review each other's results and share their interpretations. To make this possible, data needs to be aggregated first, and a part of the research workflow standardized.

"Data science can contribute to digital humanities, but that's not enough. Accelerating innovation in digital humanities through data science is our mission," Prof. Kitamoto said. He wants to provide scholars' a tool to share with others the knowledge and materials that are locked in their brains. "Anyone can curate images that they find on any IIIF-enabled websites–that's my vision for the participatory citizen science that we are trying to promote," Prof. Kitamoto said.

Discoveries at the Crossroad of the Humanities and Computer Science

One of the COHD's curation series that has drawn the most public participation is the Dataset of Edo Cooking Recipes. The COHD launched the series in 2017 with a collection of egg recipes from a cookbook "Manbo Ryori Himitsubako"

Prof. Kitamoto at the National Institute of Informatics (Chiyoda-ku, Tokyo) where he serves as associate professor

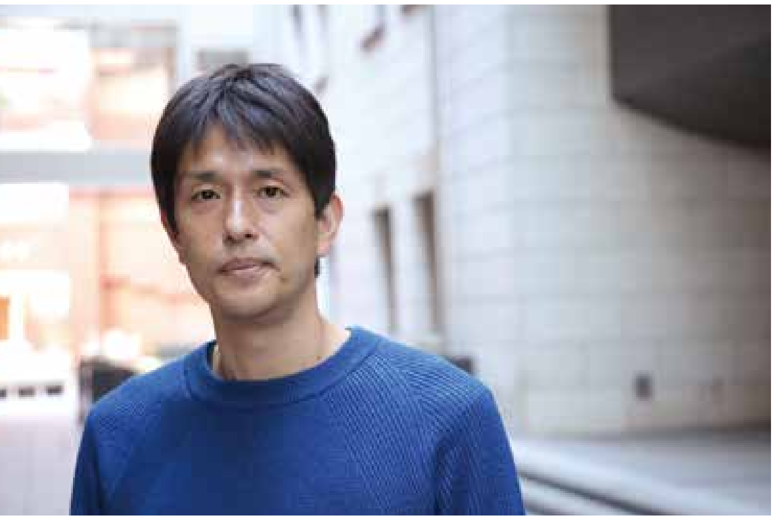
published in 1795. The egg recipes have even made their way onto a popular online recipe service, Cookpad, to reach the general public across Japan.

This experience was a lesson to learn that, just make data accessible is not enough to make the full use of it, according to Prof. Kitamoto.

“What’s interesting about humanities data is that we can study connection to the history and culture behind the data. I hope to use humanities data to not only advance computer science but also gain new understanding in the humanities,” Prof. Kitamoto said. “Computer scientists need to go beyond the research practice of putting someone’s data into a black box and get the result. When the same cutting-edge analysis tools become accessible to non-computer scientists, our field cannot grow further. We have to think about better ways to access new data, what kind of new knowledge we could gain from data, and those sorts of things. In other words, I believe that the future of our field depends on how we try to effectively and creatively handle data,” Prof. Kitamoto said.

Released on: September 26, 2018
(The Japanese version released on February 10, 2018)

Tracing the Cosmic History Back to the Big Bang



Ask the Experts: Naoki Yoshida
Professor of Department of Physics,
University of Tokyo

Dr. Naoki Yoshida is a cosmologist and astrophysicist working to solve mysteries surrounding dark matter and black holes. He has served as specially appointed professor at Kavli Institute for Physics and Mathematics of the Universe at the University of Tokyo and as professor of astrophysics at the University of Tokyo since 2014 and 2012 respectively. He received his bachelor of science in ærospace engineering and doctorate degree in science from the University of Munich and has worked at Harvard University's Department of Astronomy and at Nagoya University in Japan.

Ask the Experts: Shiro Ikeda
Professor of The Institute of
Statistical Mathematics (ISM)

Dr. Shiro Ikeda is a professor at the Institute of Statistical Mathematics. His research focuses include separation of acoustic signals with the use of "independent component analysis," a multivariate statistical analysis technique, as well as the development of processing and analysis techniques for noisy data. He also works to find out methods to improve astronomical and physical measurement through sparse modelling based on the premise that estimated target signals contain many zeros. He received his doctorate degree in engineering from the University of Tokyo in 1996.

Ask the Experts: Mikio Morii
Project Assistant Professor of
The Institute of Statistical Mathematics (ISM)

Dr. Mikio Morii serves as project assistant professor at the Institute of Statistical Mathematics. He specializes in observational astronomy.

Type Ia supernovae identified through analysis. They are compiled in data base along with the information on their luminary levels and transformations.

Modern cosmology tells us it's been 13.8 billion years since the universe emerged. Following the Big Bang, the hot and dense universe kept cooling down through continuous expansion. About 380,000 years later, the universe became transparent to radiation, enabling us to directly observe the universe. Approximately 100 million years after the Big Bang, the first generation of stars were born, and they illuminated the universe. Cosmology, the study of the origin and evolution of the universe, has risen from serendipitous discoveries that people made while gazing into space and absorbing all the information they could see, said Prof. Naoki Yoshida of Kavli Institute for the Physics and Mathematics of the Universe at the University of Tokyo. "Cosmology, from its beginning, has always been a frontier of data science," he said. Researchers today are trying to build on this cosmology-data science connection by taking advantage of the ability to collect literally astronomical amounts of observational data. Will the technology help elucidate how the universe came into existence? Three experts involved in such transdisciplinary research efforts will answer the question and explain what is happening on the cutting edge of cosmology and data science.

Key to Finding Out How the Universe Appeared

Prof. Naoki Yoshida is involved in a 5-year space observation project to use an advanced camera newly installed in the Subaru telescope to collect and analyze imaging data, which is expected to total as much as 1 petabyte in data size.

"A single image typically has several thousand galaxies captured in it. They may look like a cloud of fog at first, but the resolution is so high that you can zoom in far enough to identify individual stars. You can use the camera to capture what's happening in a vast area of space," Prof. Yoshida said. "People often think the universe is a calm place and looks pretty much the same at all times. But when you compare an image you just took from the one from the day before, you will know just how much has changed in 24 hours. In our daily analysis of the data collected with the Subaru telescope, we normally find evidence for explosions of about 100 stars a night," he said. "You can, of course, zoom into each supernova and see the details with your own eyes, but there are just too many of them to check on manually. So, we use the computer to gauge differential values between images to understand the changes occurring in the universe."

What's behind Prof. Yoshida's focus on supernova explosions is Saul Perlmutter, a 2011 Nobel Prize winner work. In 1998, the U.C. Berkeley astrophysicist used his analysis of Type Ia supernovae in the distant universe to prove that the universe continues to expand today, and that the rate of expansion is accelerating. Prof. Yoshida's research centers around Type Ia supernovae, which he said make up about a half of supernovae observed through the Subaru telescope.

"Artificial intelligence (AI) now has remarkable image

The Institute of Statistical Mathematics shares its building in Tachikawa City, Tokyo, with the National Institute of Polar Research.

analysis capabilities and can distinguish Type Ia supernovae from other types. Prof. Shiro Ikeda of the Institute of Statistical Mathematics has been instrumental in helping us develop a tool for identifying Type Ia," Prof. Yoshida said.

The question is how accurately AI can distinguish Type Ia from others, according to Prof. Shiro Ikeda of the Institute of Mathematical Statistics.

"Getting AI to accurately identify which supernovae are most likely Type Ia is our challenge," Prof. Ikeda said.

Pursuing High Accuracy in Identifying Type Ia

Type Ia supernovae are born when white dwarf stars with about 1.4 solar masses explode, and they are almost equally luminous, according to Dr. Mikio Morii of the Institute of Statistical Mathematics (ISM).

"You can estimate Type Ia supernovae's distances based on how bright they appear to our sight. You can also calculate the speeds at which they are moving away based on spectral observations," Dr. Morii said. "Based on the distances and speeds of supernovae, you can estimate how fast the universe is expanding. Accelerating expansion means the distant universe is expanding faster than originally thought."

Accurately identifying Type Ia supernovae isn't easy, however.

"In fact, 99 percent of what appear to be supernovae turn out to be something else," Prof. Ikeda said. "We compare two images taken at different times to find supernovae, but sometimes the differences aren't as clear. We select 'potential Type Ia' from an image, and that typically makes up less than 1 percent of the changes we observe from the images. We would then use a telescope to continuously observe those stars that we think may be Type Ia. After collecting enough spectral data on them, we put it together with the stars' locations and speeds to come up with a complete profile on each of them" Prof. Ikeda said. "After making great strides with accuracy, we are now able to identify 10 to 20 potential supernovae a night."

Prof. Ikeda's expertise is statistical science, not cosmology.

"When you have observation data contaminated by noise, you need to analyze data in order to understand what the information behind the noise is. Doing that is a statistical scientist's job," Prof. Ikeda said. "We call it signal processing when the data is comprised of physical signals. It often requires multiple steps using various signal processing methods to understand the original signals. I have studied such methods for astronomical data, as well, including optical light data collected with instruments like the Subaru telescope and data captured with radio telescopes. The field of signal processing is active and growing, as the advancement of measuring instruments provides us with more opportunities to collect new kinds of data," Prof. Ikeda said.

Mapping the Distribution of Dark Matter

Prof. Yoshida's 5-year observation project also focuses on gravitational lens.

"We find a number of galaxies with distorted shapes all around space. They are not physically distorted but just look distorted because of gravity that bends light-rays. The degrees of shape distortion reveal heavy objects have large quantities of some sort of invisible matter around them," Prof. Yoshida said.

This "some sort of invisible matter" is what is also known as dark matter. It is believed that the universe is made up of 4 percent chemical elements and 22 percent dark matter, with the remaining 74 percent being dark energy.

"We are investigating large areas of space to understand three dimensional distributions of dark matter, which is, in essence, a map of the universe," Prof. Yoshida said.

Scientists around the world have been studying the large-scale structure of the universe since the 1980s. But, "Just figuring out dark matter distributions based on gravitational calculations can help further theoretical models," Prof. Yoshida said. "We want to look at every distorted galaxy, including very subtly distorted ones, and compute the degrees of distortion. We can then begin to think about the nature of dark matter that makes the galaxy look the way it does through gravitational lensing. That provides basis for developing new theories and models. But, in order to map dark matter's distributions, you need to start with observations and trace them back to causes, which is a very difficult thing to do. This is another area in which we pull together the ISM's resources to support the research," he said.

The Universe Dœsn't Show You Its True Self

Separately from the 5-year observation project, Prof. Yoshida is also working on computer simulations to show how the universe began and how stars and galaxies are formed.

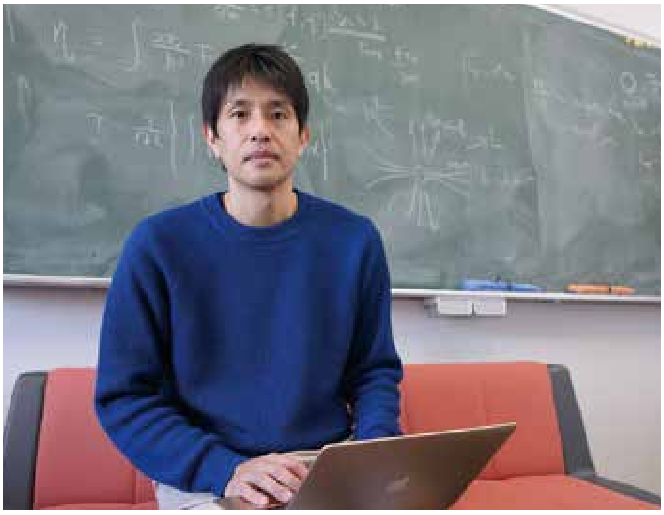
The arcs shown around the bright stars are the examples of "distortion." Strong gravity bends straight light-rays to give distant galaxies distorted appearances.

"The universe is not nice enough to show you the true picture of how its components, such as galaxies, are distributed. So, we do our best to estimate the distribution based on observations. But you can create an accurate distribution map of matter through computer simulations," Prof. Yoshida said.

"Our goal is to shed light on the evolution of the universe. We will use whatever it takes to get to the goal, including AI and super-computers," he said.

Prof. Yoshida believes that it's possible to understand what's happening around the universe using the knowledge and experiences we have on Earth.

"Saying, 'Space is nothing like here,' won't get you anywhere," he said. "Contrary to what some people might think, verified facts from our experiences on Earth provide the foundation for understanding the universe. If you mobilize utilize all the knowledge we have and still don't understand something, then we can identify what's missing. Once we've exhausted the process and realize there's a critical piece missing that we don't have, that's when a discovery happens."



Prof. Naoki Yosida at his lab at the University of Tokyo's Hongo campus

Released on: October 10, 2018
(The Japanese version released on March 12, 2018)

Data Intelligence for Sport: Put Some Smarts in Your Cheers

Ask the Experts: Yoshiyasu Tamura
Project Professor of
The Institute of Statistical Mathematics (ISM)

Dr. Tamura serves as project professor at the Institute of Statistical Mathematics. He works to develop statistical methods and studies various applications of the methods. He is also known for his research on physical random number generator, a crucial tool for simulations conducted by the ISM. He received his Ph.D. in science from Tokyo Institute of Technology. Dr. Tamura was appointed to his current position in 2018 after as associate professor and professor at the institute since 1986 and 1997 respectively.

Ask the Experts: Fumitake Sakaori
Associate Professor of
Chuo University

Dr. Sakaori serves as associate professor for the Department of Mathematics, Faculty of Science and Engineering, at Chuo University and as visiting associate professor at the Institute of Mathematical Statistics. He specializes in statistical modeling, computational statistics, statistical science in sports, and statistical education. Dr. Sakaori was appointed to his current position at Chuo University after working as assistant professor for College of Sociology at Rikkyo University and as adjunct professor at Chuo University's Faculty of Science and Engineering.

We have seen the scenario come true before: A cash-strapped professional baseball team rises up to win its league championship, out of left field. Drastic performance improvements of sport teams are often the results of game strategies developed by utilizing data intelligence. In baseball, "sabermetrics"–statistical analyses of players' performances and game activities–began garnering attention in the 1980s. Inspired by how the American Major League Baseball teams quickly adapted sabermetrics as the key component of game-plan strategy, the Japan Statistical Society established the Sport Data Subcommittee in 2009 to spearhead sport applications of data science in the country. Having recently celebrated the seventh anniversary of the group's ever-growing flagship annual event, "Sport Data Analytics Competition," the subcommittee leaders believe data analytics are destined to serve as the fundamental tools in sport teams' pursuit of victory. Prof. Yoshiyasu Tamura of the Institute of Statistical Mathematics, Prof. Fumitake Sakaori of Chuo University and Prof. Akinobu Takeuchi of Jissen Women's University–three of the founding members of the subcommittee and the contest–will share their takes on the future of the application of data science in sport and what it means for the talents that will be required to help develop successful game strategies.

Ask the Experts: Akinobu Takeuchi
Professor of
Jissen Women's University

Dr. Takeuchi is professor for the Faculty of Humanities and Social Sciences at Jissen Women's University and the director of the university's Information Center. He specializes in statistical science, behavior metrics, and statistical education, including data science education. As the chairperson of the Special Committee of Statistical Education, Dr. Takeuchi has long worked to develop educational methods for fostering statistical thinking.

Prospective Data Scientist Turned Professional Baseball Player

On March 19, 2018, dozens of data scientists from across the country, both professionals and students, descended on the Research Organization of Information Systems (ROIS)'s the Institute of Statistical Mathmatics' headquarters in the suburban Tokyo city of Tachikawa for the 7th annual Sports Data Analysis Competition award ceremony. The event highlighted the work of a grand prize, first prize and honorable mention winners in the highly competitive contest, which began in June 2017 with a call for submissions, followed by a September deadline and announcement of the awardees in January. The exponential growth of the contest's popularity over the past years speaks for the surging interest in use of data analytics in sport, said Prof. Yoshiyasu Tamura of the Institute of Mathematical Science, a baseball enthusiast who used to play the sport for his university as a catcher.

"I love all kinds of sports, and that's how I became interested in applying statistics to sports," Prof. Tamura said, looking back on his college days. "When we began the competition, very few people had heard of the word, 'data science,' and we didn't even have the Japan Data Scientist Society," Prof.

The results of the 7th annual Sports Data Analysis Competition are available at https://estat.sci.kagoshima-u.ac.jp/sports/ (in Japanese).

Tamura said. “A lot of people are sports fans like me. It makes sports an ideal subject for practicing analytics.”

The success of the Sports Data Analysis Competition has proven him right. The 2017 contest drew submissions from as many as 61 teams for three sports categories–baseball, soccer and basketball–each of which is broken down into three subcategories of “Analytics” and “Infographics” for adults and “Secondary Education” for middle and high schoolers.

Entries ran the gamut in subject matter from “Topological data analysis for the evaluation of the robustness of the defense system,” submission by a Keio University team, to “Analysis of basketball scores” entered by a Chuo University team, which won the grand prizes in the analytics and infographics categories respectively.

“When I think about this year’s entries, there was more quality research that was soccer-related than baseball-related,” Prof. Tamura said. “In baseball, you only have pitch-by-pitch data. In soccer, on the other hand, you not only have data on an entire play but also ball tracking and touch-by-touch data, allowing you to synthesize them all for analysis. The quality of soccer data is really improving,” Prof. Tamura said.

The competition organizers are trying to offer participants more sport choices to choose from, as well, which was the reason behind the addition of the rugby category for the 2016 contest.

“Statistics can be difficult to analyze, but sports have the magical power to get people excited about analyzing data,” Prof. Tamura said. “I know one university student, who participated in the competition twice, went on to become a professional baseball player. I was stoked to find out about that,” he said. “Every professional baseball team nowadays is on the hunt for data analysis talent.”

Climbing the Learning Curb with Sport Data Analysis

Contest judges first screen the submissions based on how attractive the analysis themes are.

“If there is a surprising result or highly practical use for the research, we would add points for that, too,” said Prof. Fumitake Sakaori of Chuo University, who has led the competition long with Prof. Tamura since its inception. “In other words, we look at whether the analysis is warranted.”

The second most important criteria are whether the teams used appropriate techniques for analysis as well as their ingenuity in the ways they conducted their analysis, according to Prof. Sakaori. Originality and novelty count, he said.

“Teams that show originality in data analysis and groups that try novel processing techniques to produce their own data instead of using the variables given by the competition organizers can earn extra points, as well,” Prof. Sakaori said.

“The great thing about sports is that those analyzing the data easily understand what it is that they are trying to get out of their analysis. Thus, sports data makes for excellent educational material, especially when one is familiar with whichever sports in the subject of analysis.” Prof. Sakaori said.

He explained that the Sport Data Subcommittee’s single biggest goal in organizing the competition was to encourage people to make use of data and grow their ability to think in terms of mathematical statistics.

“The competition program has been very effective in that light,” Prof. Sakaori said. “I can see contestants’ skills are improved every time they come back for another competition. From the angles they take to their analysis to their objectives, everything gets better. I suppose looking at excellent work done by other teams inspires you and helps you develop statistical thinking.”

In the end, you have to ask yourself for whose benefit you are analyzing data–whether it’s for the players, for the coaches or for the team owners–or whether you are doing it to enhance the experience of watching games on TV, Prof. Sakaori said.

“Once you answer these questions, it becomes clear how you should present your analysis,” he said.

Looking Beyond 2020: Data Analysis May Be Sports’ Future

In the Sports Data Analysis Competition, participants get to work with streamlined datasets provided to them. In real life, however, it is difficult to find well-organized quality data, and in some sports, you couldn’t find any data at all, according to Prof. Sakaori.

“That creates a challenge for people who are trying to tap into the power of data analysis to contribute to the sports industry,” Prof. Sakaori said. “Another challenge is how to produce analytical data from things like video footage of sports tournaments. You can take data from video of a judo tournament, for example, but that’s unstructured data and you wouldn’t necessarily know how to approach it. We must advance our knowledge in not only statistical science but also informatics in order to deal with these issues,” he said.

Some countries have all sorts of professional sports leagues, including basketball leagues, soccer leagues and American football leagues. In these countries, there are many science magazines focused on sports statistics, and lots of research being conducted on the topic, as well, according to Prof. Sakaori.

"In Japan, we are now seeing more sports consulting businesses being established to work with professional teams and strategize how to boost the industry for the post 2020 Olympics era," Prof. Sakaori said. "Since the Japan Sports Analysts Association (JSAA) has also been established in 2014, a number of conferences have been organized, and they are growing larger every year, offering opportunities for sports analysts to compare notes. If this is any indication, I predict sports games 10 years from now will look much different from the ones we see today."

Prof. Sakaori explains data at the 7th annual Sports Data Analysis Competition award ceremony. The infographics category of the 7th competition drew fewer submissions than expected. Prof. Sakaori says he hopes prospective contestants to give a lot of thought to what they want to communicate through infographics.

The results of the 7th annual Sports Data Analysis Competition for the Secondary Education category is available at https://estat.sci.kagoshima-u.ac.jp/cse/sports.htm (in Japanese).

Everyone Needs to Be Able to Talk Data

The 7th annual Sports Data Analysis Competition marked its fifth year since opening the contest to middle- and high-schoolers and drew as many as 65 submissions to the Secondary Education category. Prof. Akinobu Takeuchi of Jissen Women's University, who has overseen the category since its beginning, said he has been impressed by the quality of the students' work.

"People from around the world often look at the Japanese as being highly skilled in mathematics. But statistics reveal that many other countries, particularly China, the U.S., South Korea, New Zealand, England and Australia, provide much more advanced math education than Japan does," Prof. Takeuchi said.

There is a movement to revamp math curriculum in Japanese secondary education with an eye toward providing all students with the necessary statistical skills to live in the data age. Prof. Takeuchi has actively been involved in that initiative in recent years.

"Until recently, middle and high schools in Japan weren't doing enough to teach the concept of an "average" and how the "median" value and "mode" have statistical significance. It makes you wonder how this will prepare the students for their future–if they will become capable of joining their counterparts from around the world even for a casual business conversation," Prof. Takeuchi said.

He said that the sports data used in the competition are "pretty clean and almost normally distributed." Data that statistical scientists have to work with in real life, including social survey data, is never that clean, he said, but having neat data helps contestants explore and discover how to effectively use data to draw useful analysis from it.

"At the end of the day, you want your sports team to win. But, just because they won, it's not necessarily clear why they won or what worked the most. We have to figure that out scientifically. And we want the contestants to learn the thought process required in statistical science," Prof. Takeuchi said. "If I take the basketball as an example, we want them to analyze why the team has a low shooting rate, what winning teams or losing teams have in common and other factors by using various analysis techniques. It is important to be able to realize a new technique could shed light on causal relationships underlying what they've found."

This year's entry from a team from Kannonji Daiichi High School in Kagawa prefecture epitomized such learning, according to Prof. Takeuchi. The team conducted analysis to figure out ways to strengthen a local basketball team. The students created a diagram for their poster that summarized the causes and effects of game results. They conducted a principal component analysis and verified the analysis method, as well. The poster, which proposes a new strategy for the basketball team, won the grand prize in the category.

"Regardless of whether you are interested in majoring science or want to study the humanities and become a marketing professional, data science skills provide the foundation for all. This competition's purpose is to raise the next generations who are fluent in the language of data."

Released on: November 12, 2018
(The Japanese version released on April 12, 2018)

Changing Patients' Lives with Data Science

Ask the Experts: Shin'ichi Satoh
Professor of
National Institute of Informatics (NII)

Professor at the National Institute of Informatics since 2004 after serving as associate professor at the institute since 2000. Expert in image and video analysis-based information search and knowledge discovery. Previously worked at Carnegie Mellon University as visiting research associate from 1995 to 1997. Holds a Ph.D and undergraduate degrees in engineering from the University of Tokyo.

Ask the Experts: Yoichi M. Ito
Professor of The Institute of
Statistical Mathematics (ISM)

Professor at the Institute of Statistical Mathematics. Specializes in biostatistics with a focus in genomic and statistical analysis and provides consultation on the topics. Has served as an expert member on the PMDA, a reviewing board for new drugs, since 2009. More recently, working to analyze clinical trial data through the lens of efficiency to develop ways to improve data management. Holds a Ph.D. in health sciences from the University of Tokyo.

Saving patients' lives isn't just the responsibility of medical and healthcare professionals any longer. It takes a village to provide the best care possible, and playing increasingly important and larger roles in it are data scientists and people in the associated fields. This trend precipitated the recent openings of the two national data research centers–one at the National Institute of Informatics (NII) and the other at the Institute of Statistical Mathematics (ISM). The Research Center for Medical Bigdata at the NII opened in November 2017, followed by the Research Center for Medical and Health Data Science at the ISM's opening in April 2018. Both the NII and the ISM are part of the Research Organization of Information Systems (ROIS). The two new research centers work to muster all the information technology (IT) and statistical knowledge and resources available in Japan while promoting interdisciplinary collaborative research projects as members of the Inter-University Research Institute Corporation. Well-managed data infrastructures can propel scientific discoveries and innovations, and so does superior analytical capacity, such as that of artificial intelligence (AI). Now that the newly opened research centers aim to deliver both, how will that help drive forward Japan's medical research and health care? The directors of the two centers will share their visions for future applications of data science in medicine.

Ask the Experts: Hisashi Noma
Associate Professor of
The Institute of Statistical Mathematics (ISM)

Associate professor at the Institute of Statistical Mathematics. Appointed to the current position after studying biostatistics at the Graduate School of Medicine, Kyoto University and serving as the ISM's assistant professor. An expert in the fields of medical statistics and public health, he hopes to "contribute to the advancement of medicine and health care through research and education."

Bringing the Best of the Best in the IT World to Help Physicians

The National Institute of Informatics (NII), which is home to the newly opened The Research Center for Medical Bigdata, is Japan's only general research organization solely focused on information technology (IT). The NII became the Japan Agency for Medical Research and Development (AMED)'s partner organization in 2017 and has since worked with four academic societies–the Japan Gastroenterological Endoscopy Society, the Japanese Society of Pathology, the Japan Radiological Society, and the Japanese Ophthalmological Society–to develop the IT infrastructure for the medical and health care communities. The collaborative efforts include image analysis, deep learning and work toward the development of AI.

The Research Center for Medical Bigdata now leads these efforts. The center has about 10 research team members, some of whom are enlisted from such outside organizations as the University of Tokyo, Nagoya University and Kyushu University. The members also include some young post-doctoral researchers. They are among the brightest in their own fields, but most of them don't have any experience in analyzing medical images.

"I, for one, am a researcher specializing in image analysis and deep learning and had never dealt with medical images before," said Prof. Shin'ichi Satoh of the NII, who serves as the director of the Research Center for Medical Bigdata. "The thing is, we can analyze medical images. Though we don't know what we are looking at, the artificial neural network that we trained, which is a type of AI, does know what to do," he said.

High-performance image-recognition systems aren't foolproof. "If you add carefully crafted 'noise' to images of pandas and feed them to a machine, for example, 99.3 percent of the time, the system will tell you they are the images of a gibbon," Prof. Sato says. "We need to test the machines from various angles to understand how they process something out of the script."

Until recently, scientists in the field of medical image analysis were trying to create computer programs for data processing that reflect their medical knowledge, according to Prof. Sato.

"There has been a 'shift in the tide,' if you will, as we became capable of feeding an astronomical amount of data to a machine to make some sense of it," he said. "It's been only a year and several months since our collaborative research began, but we are already seeing a lot of successes with the use of machines. For example, very basic artificial neural networks can quite accurately diagnose diabetes and glaucoma, particularly when using fundus images as a diagnostic tool. These diagnoses are actually challenging ones for physicians to make," he said.

A Turning Point in the 'Game of Computer Image Recognition's

Prof. Sato pointed out there have been three major "waves" that pushed forward the research on artificial neural networks over the years.

"The first wave was the emergence of a neural network that mimicked that of the human brain. This was technically a single-layer network capable of learning how to use randomized linear functions. It proved that machines can learn from inputs and became the framework for machine-learning. The second wave was the arrival of so-called neuro-fuzzy. This was a little more complex network with two to three deep-learning layers," Prof. Sato said.

"Then, in 2012, a revolutionary algorithm mimicking the human brain's neuron network became available. Its ability to learn and think far exceeded that of any other algorithm that existed at the time. This spurred interest in leveraging the technological advancement and availability of big data to make discoveries. This is the third and latest wave in the history of neural network inventions," he said.

The revolutionary computer vision algorithm was built by University of Toronto professor Geoffrey Hinton–who is known in the research community as the "Godfather of AI"–and his students for a competition in which they participated. The team gave the large, deep convolutional neural network 1.2 million images to train it to recognize 1,000 different

types of objects.

"The team's research paper gave researchers an understanding of how much data a machine needs to do its job and at what level of accuracy. It became a turning point in the game of computer image recognition," Prof. Sato said.

"Before this, what objects a machine should recognize and how it should be done were all determined by humans. What sets Prof. Hinton's neural network apart from the rest was that it was designed to teach itself from scratch what to do and how," Prof. Sato said. "When you train this type of network by giving a huge number of images containing 1,000 different objects, it does a superb job recognizing objects in newly introduced images. Unfortunately, we don't know what the machine learned, but it clearly can carry out very efficiently what we humans used to rack our brains to do."

Prof. Hinton's algorithm had eight learning layers. Nowadays, deep-learning machines with 100 to 200 layers are commonplace, according to Prof. Sato.

The theoretical analyses that provide the foundation of deep learning have remained unchanged for the past three decades, Prof. Sato said. But, what you can do with artificial neural networks, especially with computer vision recognition, has drastically improved over the years, and researchers have now had experiences to better understand the possibilities as well as the limits that artificial neural networks present, he said.

At the Research Center for Medical Bigdata at the NII, researchers are working toward finding efficient ways to use computer image recognition technologies for analysis of medical big data.

Making Doctors' Lives Easier

Prof. Sato believes image recognition algorithms can help reduce physicians' workload. To achieve this goal, his team is compiling "correct diagnostic interpretations" of image data that can be used to train the machines.

"If you were analyzing image data to correctly recognize cats and dogs, anyone could do it. But medical analysis requires physicians' expert knowledge. Data has to be looked at by physicians, and they need to tell you what the right interpretation is. This defeats the purpose of making physicians' lives easier," Prof. Sato said. "So, we use self-learning algorithms to narrow down the scope of data that requires physicians' judgement and request doctors to attach correct interpretations to it. This considerably reduces physicians' workload. But, I think we can and should make the algorithm even more physician-friendly," he said.

Algorithms could also assist physicians in other ways to improve the quality of medical care.

"In the near future, we may have an image analysis system designed to provide support to physicians. The system could prevent physicians from overlooking important information or offer an independent second opinion. It could also diagnose minor illnesses, so the physicians can use their time to focus on other illnesses that are more difficult to diagnose," Prof. Sato said. "It's our dream to develop an advanced AI algorithm that has the judgement of experienced physicians. If we create that, we can then make it available in remote places like islands. It's a program, and so you could easily replicate it, as well," he said.

Prof. Yoichi Ito (left) and Dr. Hisashi Noma, who serve as the director and vice director of the Research Center for Medical and Health Data Science at the ISM respectively.

Fostering Data Scientists in University Medical Departments

In the meantime, the Research Center for Medical and Health Data Science, which recently opened at the Institute of Statistical Mathematics (ISM), aims to bring together medical statistics experts and build different types of intelligence infrastructures to prepare the country for medical and health care in the coming data age.

“More than 60 universities across Japan have medical departments, and they all critically lack biostatisticians. The center’s missions include fostering the next generation of data scientists with the knowledge and skills required to assist the medical community, as well as providing training for literacy in medical statistics,” said Prof. Yoichi M. Ito of the ISM, who serves as the director of the center.

The center offers four courses a year on education of medical statistics. Five publicly accessible seminars have already begun as part of these courses, and a symposium organized to celebrate the center’s inception was also held on May 28, 2018.

“The foundation of this center would help raise the level of enthusiasm for data science in the medical and healthcare community, and the change was palpable,” said Dr. Hisashi Noma of the ISM, who serves as vice director of the center and directed the symposium’s operation.

Prior to the center’s opening, the ISM called for the formation of the Medical and Health Data Science Research Network to promote statistical education. More than 70 institutions, including universities and hospitals, quickly answered the call, Dr. Noma said.

Rigorous Evaluation of Clinical Trials from Statistical Science’s Standpoint

Evaluation of clinical trials for new drugs is another area where Prof. Ito believes the Research Center for Medical and Health Data Science can make contributions.

“As you know, statistics has always played a critical role in life science, particularly pharmaceuticals. I mean, statistics is an essential tool for understanding the effectiveness and safety of a new drug when evaluating the result of a clinical,” said Prof. Ito, an expert of clinical trial evaluation who has years of experience in helping design clinical researches as a statistical consultant.

“The human body has biologically complex systems and exhibits very different responses to drugs than other species do. Over the years, we have seen human subjects in some clinical trials experience totally unexpected side effects to drugs that had proved safe in mice, dogs, and even monkeys that are biologically more similar to humans. To understand whether a new drug has any therapeutic benefit or if it’s safe for humans from the result of clinical trials, we must apply statistical science to come up with an exact margin of error and scientifically evaluate the data. This is where statistical skills come in,” Prof. Ito said.

“Design methods for clinical trials are also improving every day, and so we need to keep pace with new methodologies for statistical analysis. This is a highly specialized field in which every step is intentional and calculated–which makes a sharp contrast with analysis of big data,” Prof. Ito said.

Accurately Estimating Harmful Environmental Factors

Epidemiology, the study of distributions and determinants of human health and diseases, is another area in which statistical science is making a big difference. The history of epidemiology goes back to the 19th Century England when scientists began researching what caused the spread of cholera and other infectious diseases and how to prevent them. Since then, it has evolved into the study of medical care and health in general as disease patterns have diversified in modern society. More specifically, epidemiologists refer to those who make scientific assessments of the health impact of such environmental factors as cigarette smoke, hormone-disrupting chemicals and air pollutants, using observational studies of humans as their basis, according to Dr. Noma, who has

Prof. Ito says he often fell ill as a child and wanted to become a doctor to help people like him. Now in his ninth year of serving as an expert member of the Pharmaceutical and Medical Devices Agency, Prof. Ito reviews a lot of new drug application from an outsider's point of view. "This is an important role through which I can contribute to improving the quality of new drug reviews," he says.

The Research Center for Medical and Health Data Science has recently published several new "meta-analysis" methodological papers. Meta-analysis refers to synthesis analyses of systematic evidence from past clinical trials to investigate the effectiveness and safety of medical technologies, according to Dr. Noma.

worked on epidemiological research projects since college. "The problem is that we have no way of controlling those harmful environmental factors like you would in normal scientific experiments. We rely on statistical methodology to precisely determine how and from which subjects we should collect data and whether other factors are involved," he said.

Epidemiological experts particularly try to look at data through the lens of statistical science to minimize the effect of biases.

"Measuring biases is very difficult. It's just not possible to set 'perfect' questions that produce no biases in evaluation," Dr. Noma said. "But, because you are evaluating public health risks, even though you know what you have includes a bias, you have to take action in response to it to protect the public. Then you again need the help of statistical science to figure out what an effective action would be."

Epidemiology began a long time ago, way before statistical science did. In recent years, the methodologies used in epidemiology started becoming more refined. Methodologies didn't originate in clinical trials, which is my specialty, but the fields of epidemiological statistics and life science statistics, such as what you obtain in clinical trials, are crossing over, spurring advancement of methodologies," Prof. Ito said.

Building New Methodologies from the Base

The Research Center for Medical and Health Data Science has six ongoing research projects, and one of them is called, the "Project to develop methodologies for clinical researches and evidence synthesis." The project brings together the center's dual interest in developing novel methodologies and promoting big data analysis to make "precision medicine" available in daily medical practices. Efforts to advance research on precision medicine are taking place around the world, including in the United States where former President Barack Obama announced the Precision Medicine Initiative during his 2015 State of the Union address.

"For example, thalidomide is infamous for bringing on horrible side effects in patients worldwide during the 1950s and 1960s. But, in 1999, a group of scientists in the U.S. proved that the drug that prevents angiogenesis, can effectively be used to treat myeloma, resulting in the approval as an anti-cancer agent in Japan in 2008. Through data analyses that use data science methodology, researchers are now finding out that the nature and degrees of a drug's effects vary from patient to patient," Prof. Noma said.

"Let's say you have cancerous cells removed from a patient. You measure all sorts of things for comprehensive analysis of genetic and molecular-level information, and you will end up with more than several million dimensions of data. This is what you call omics data. Omics data has become available, partly thanks to the considerable advancement of measurement techniques in the past two decades," Prof. Noma said. "Using our new statistical methodologies to analyze patients' omics data in connection with thalidomide's effects, for example, it becomes clear that those who were treated with the drug have better prognoses than patients who didn't take the drug. In addition, thalidomide affects individual patients differently, depending on whether their molecular cells show certain gene expression patterns. By applying newer methodologies to the analysis of omics data, you can gain such detailed insights," he said.

"With research projects beginning to leverage big data in more diverse ways, methodology for data analysis is needing to change and diversify, as well. The Research Center for Medical and Health Data Science strives for fundamental research to help fulfill this demand, conducting research on such topics as the use of machine learning technology for exploratory analysis of genetic data, as well as the theoretical development of classical statistics," Prof. Ito said. "It is our highest priority to establish an analysis model to offer patients accurate prognoses, rather than to get a complete picture of a group of patients. This is the most urgent task for us. It's also an intriguing challenge for us to tackle."

Released on: December 18, 2018
(The Japanese version released on May 10, 2018)

What's on the Horizon in the Changing World of Big Data?

Ask the Experts: Masaru Kitsuregawa
Director of National Institute of Informatics (NII)

Since 2013, Dr. Kitsuregawa has served as the director of the National Institute of Informatics, the Research Organization of Information and Systems, and professor at the Institute of Industrial Science at the University of Tokyo. As an expert in database engineering, he leads some of the most prominent informatics research projects in Japan and has developed a high-speed database engine, using "out-of-order execution schemes." He is known for operating the mammoth Global Environmental Database. The 1983 graduate of the University of Tokyo holds a Ph.D. in engineering from the same university. He previously served as vice chairman of the Information Processing Society of Japan and 23rd chairman of the Informatics Committee of the Science Council of Japan.

Ask the Experts: Tomoyuki Higuchi
Director of The Institute of Statistical Mathematics (ISM)

Since 2011, Dr. Higuchi has served as the director of the Institute of Statistical Mathematics, the Research Organization of Information and Systems. An expert of Bayesian modeling, he is known for statistical modeling derived from real-world problems as well as research on data assimilation that combines mass data and theories behind simulative calculations. He leads various integrated research and works to foster the next generation of researchers. He serves on the ROIS's board of directors and as an advisor to the Japan DataScientist Society. He received his bachelor's degree in 1984 and his Ph.D. in 1989, both from the University of Tokyo.

We are living in the era of big data. Easy access to colossal amounts of data being constantly gathered and compiled on virtually every topic imaginable is enabling businesses and research institutions, such as universities, to gain insights that would have seemed impossibly far-fetched only a few years ago. As artificial intelligence (AI) becomes more sophisticated in its ability to autonomously make data-driven decisions, many of us are asking ourselves how soon we may reach the singularity — the tipping point where computer intelligence exceeds that of humans. This causes us to pause and examine in which directions data science is advancing and what kind of research we need to conduct to help solve complex problems. The directors of the National Institute of Informatics (NII) and the Institute of Statistical Mathematics (ISM) – which are both part of the Research Organization of Information and Systems (ROIS) – recently sat down to talk about all these issues and share their outlooks on the future of big data with the Science Report's readers. Here are their perspectives and the latest on the leading-edge technology and research.

Creating Big Data from Small Data

Data analysis techniques and AI technologies are constantly evolving and advancing. Where are we at in this journey, and what type of research projects are commanding the science community's attention?

Kitsuregawa: The most interesting challenge we are tasked to tackle is the question of how to make small data big. For example, typhoons hit Japan typically 10 times a year, and this means we only get 100 incident data to learn from over 10 years. The question is, how we can draw the information that we need from the data on these 100 incidents? In other words, how do we data scientists do what we do when data is intrinsically lacking in certain fields? That's the biggest challenge facing us right now.

Higuchi: Those are so-called rare events. They spur innovations and help make risk analysis more accurate. In mathematical statistics, researchers have traditionally used the "experimental design" method to prove a hypothesis through a limited number of experiments. But today's science calls for an even more intentional system–a system that merges experiments, measurements and observations into a single process.

Kitsuregawa: Prof. Takaaki Kajita of the University of Tokyo, for instance, had very little experimental data, but his interpretation of the data has led to his discovery that neutrinos have mass, for which he received a Nobel Prize in 2015. This "interpretation" was a very sophisticated one and on a completely different level than AI's pattern recognition and learning. But because Prof. Kajita was studying a phenomenon that occurs in accordance with the rules of physics, it was possible for him to trace it, step by step, to its causation. In most cases, you can't do the same with what we see happening in this world. Take rare diseases as an example. It's an area that we've been interested in, because unlike diabetes and hypertension from which millions of people suffer, we may only find one patient in Japan and another in England who have a certain rare disease. So, how do we go about finding a cure for a rare disease like that? Tackling such a problem requires us to bring together all talents and expertise and try different ideas.

Higuchi: In material data science, we may be able to

determine the molecular bond and structure that gives substance a certain feature, but it is difficult to actually create that substance. This is the bottleneck in the practical application of advanced material data science. Researchers are experimenting with different approaches, such as simulating an experiment and searching for a pattern among its results. But when it comes to figuring out the process of creating a new material, experience still remains a more reliable tool than data science.

Kitsuregawa: In other words, just because a certain calculation method worked in identifying a material's structure, that doesn't tell you how to extend the method for finding a way to create the material. Also in natural language processing, the machine's ability to use big data for deep-learning is still limited. We need to figure out our next "card" that gets us through this bottleneck. In that sense, we are in a transitional phase.

Getting Through Data Science's 'Bottleneck'

Data science is becoming an integral part of research in a wide range of fields from earth science to nanotechnology to health care to material development. While industry puts high hopes on data science as the key to innovations, directors Kitsuregawa and Higuchi say data science still has a long way to go before it can solve more complex problems.

Kitsuregawa: Science can effectively solve various problems, but there are certain areas where science can't be of help, as I mentioned earlier. I believe we, the scientists, ought to academically identify those limitations, as well.

Higuchi: That's right. I often explain this limitation issue through the concepts of interpolation and extrapolation. In data science, we basically make predictions within the bounds of existing data. That's interpolation. Extrapolation is to go beyond the bounds and make predictions. Interpolation techniques have grown tremendously sophisticated over the years, contributing to the advancement of AI. The question now is how we are going to develop extrapolation techniques to provide solutions to problems. This is where we hit the brick wall. Another limitation of data science, currently, is that while it can show how things are correlated, it's nearly impossible to demonstrate if there's any causal relationship among them.

Kitsuregawa: Yeah, it's impossible.

Higuchi: Right. I mean, you extrapolate data to predict the unpredictable, and it's not possible yet to pinpoint exactly when and where an earthquake is going to happen, for example. Not, at least, until data science advances a little more and we have far better simulation models. The good news is that we are becoming better at quantitative risk analysis by making good use of high-speed computation and abundant data collected through observation networks.

The Language of Data as Part of Basic Education

The big data era is here to stay, and data technology is expected to grow more sophisticated. So, how should we equip our children with the knowledge necessary to live in data society?

Kitsuregawa: Science and technology is advancing so much faster nowadays than in the past, following the "law of accelerating returns." Eventually, it will reach the "singularity." The biggest conundrum will be people's slowness in digesting the fact and adapting to the new reality in everyday life. Just think about how long it may take to reform the Copyright Act to make it suitable for the internet era. You may adopt new technology to make society a less confusing place, but if people don't understand its benefits, society would only grow more chaotic. I just wonder if there is a way to expedite the human recognition process.

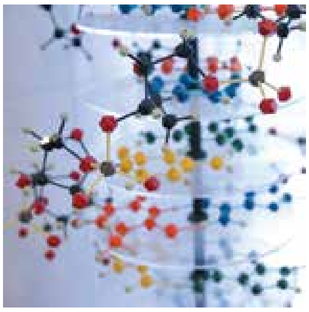
Looking back on human history, a drastic transformation or the collapse of a culture has always occurred in its very final moment–equivalent of one last second versus an entire year. In hindsight, you can see there is a sequence leading up to such big changes, which is usually triggered by an unforeseen, peculiar human act, which then develops into a war, causing fear among the population. If this is any indication, if only technologies continue to advance without adequate time for people to appreciate what's happening, it will sooner or later trigger major changes to our society. That's my concern.

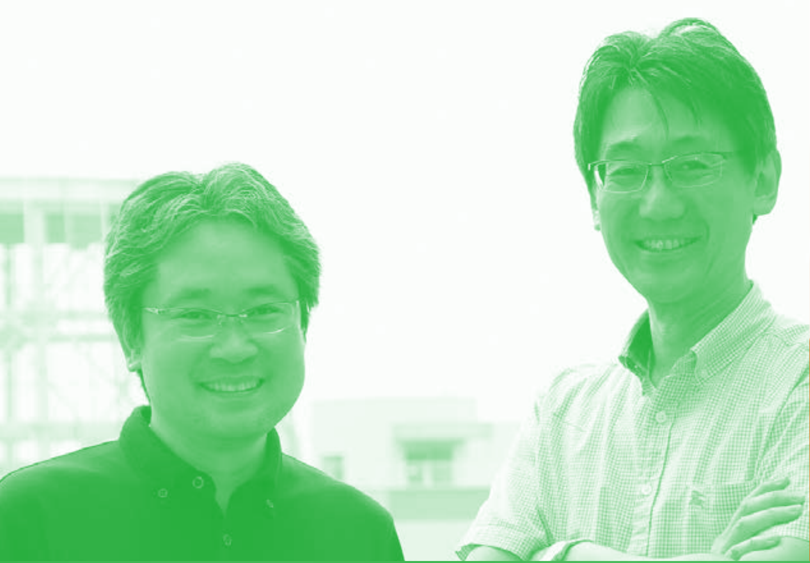
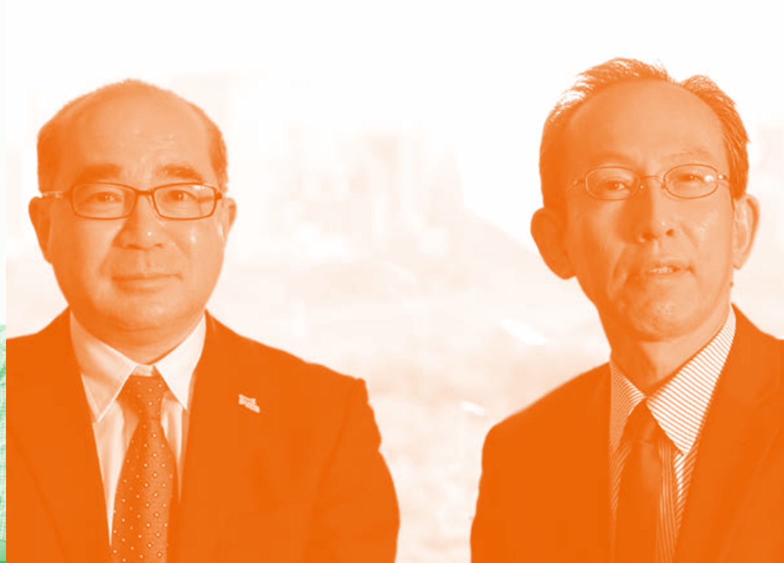
Higuchi: I feel the same way. This leads to the question of how we should educate the new generation. By the way, what are your thoughts on programming education in secondary schools?

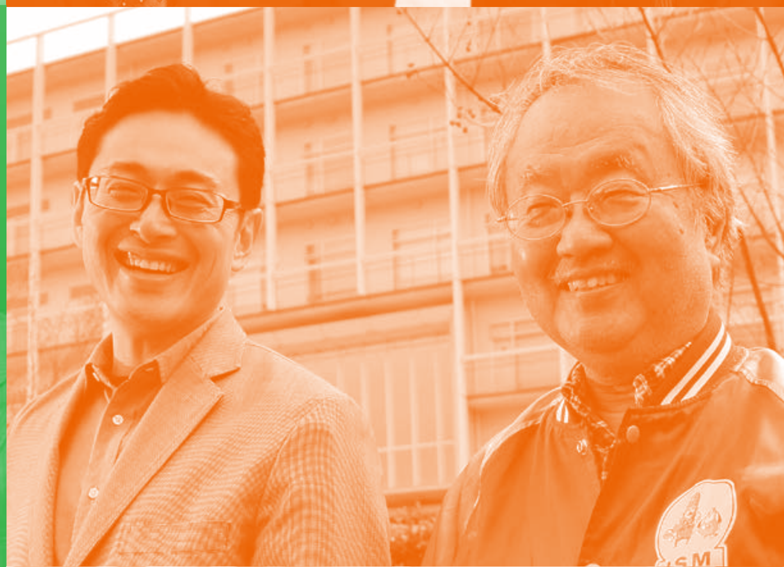
Kitsuregawa: In Japanese schools, students learn Japanese first, English some years later and another language in college. I think the education of programming language should be given as much weight as any non-Japanese language, if not more. By having good English skills, you will be able to communicate with people from around the world and understand them, and that enriches your life. Programming language allows you to turn your idea into a tangible "thing." In every IT course I teach, I always tell my students to bring something manmade to the first class. You will realize pretty much every product has a computer in it. This goes to show you that, without some knowledge of programming language, you won't be able to make anything, and you may even have difficulty striking up a casual conversation with someone. In that regard, programming language should be part of the basic education for everybody. It is crucial that people begin learning how to make things at young ages.

Higuchi: We are also surrounded by data in today's society. I hope future generations will grow up analyzing and interpreting data themselves to improve their daily lives. At the ISM, we have a program for secondary school students called ISM Data Science High School. It may sound like a place that teaches complex machine learning and analytical methods, but that's not our focus. We want to provide the students with learning opportunities to figure out on their own that how they select a problem to tackle, and what angle they take to solve it, matter the most in data science.

Released on: January 21, 2019
(The Japanese version released on June 11, 2018)










JISSEn


大学共同利用機関法人
情報・システム研究機構
Research Organization of Information and Systems

ISBN978-4-909638-09-0 C0040