National Institute of Informatics News ISSN 1883-1966 (Print) ISSN 1884-0817 (Online)

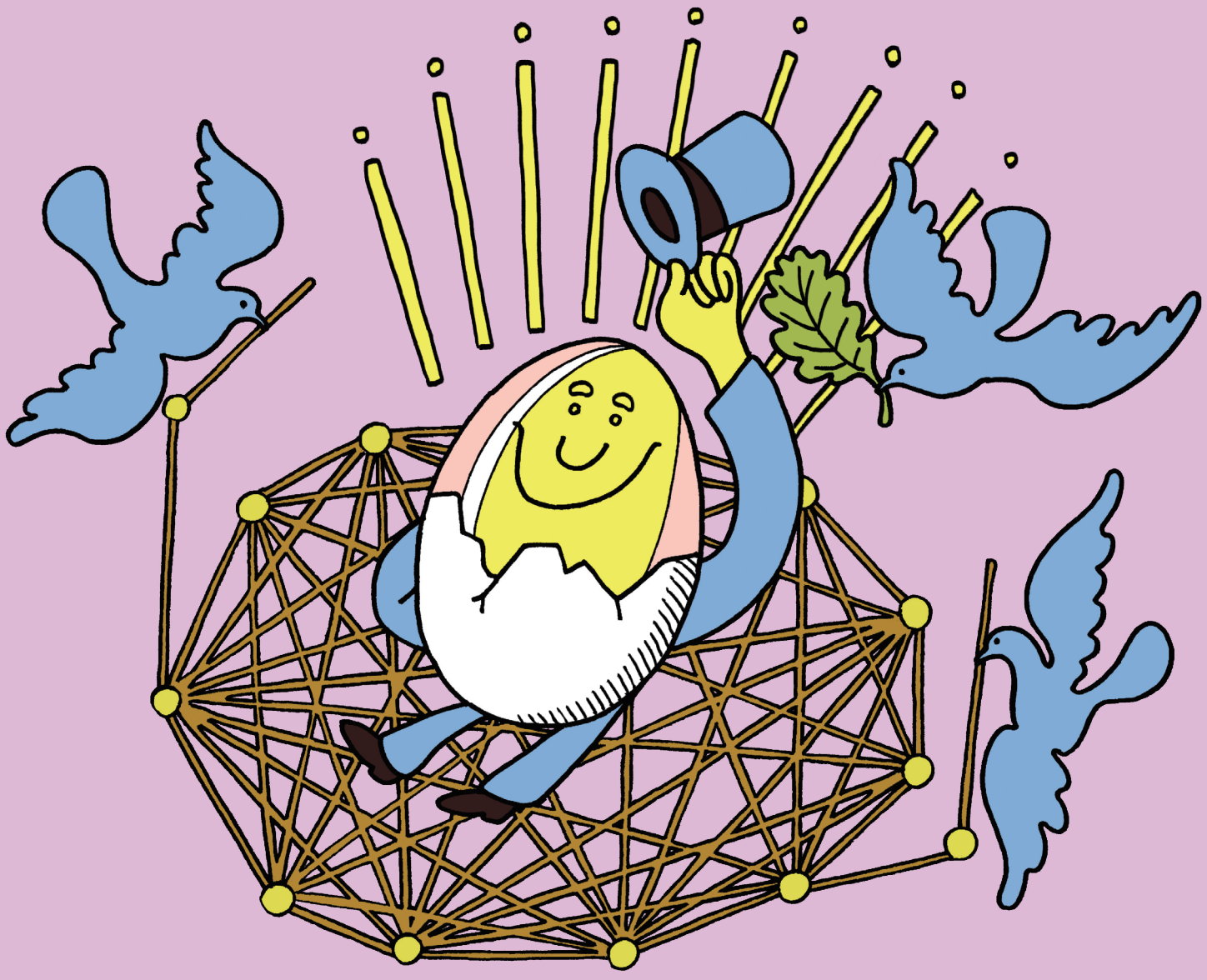
NII Today

93
Dec. 2021
National Institute of Informatics News

Feature

Kashiwa Annex, a New Knowledge Hub

Expanding academic information infrastructure services



This English translated version NII Today corresponds to No. 93 of the Japanese edition

www.nii.ac.jp/en/about/publications/today/

Feature

Kashiwa Annex, a New Knowledge Hub

Expanding academic information infrastructure services

In March 2021, the National Institute of Informatics (NII) opened the Kashiwa Annex (Kashiwa City, Chiba Prefecture) on The University of Tokyo's (UTokyo's) Kashiwa II Campus, with the aim of expanding its academic information infrastructure services. Equipped with the Science Information Network (SINET) and facilities for academic information infrastructure services, Kashiwa Annex will also serve to enhance research collaboration with UTokyo.
mdx is a new "platform for the data-driven future," introduced by UTokyo. NII will also participate in mdx as an operating institution. The NII Kashiwa Annex will play a role as a next-generation academic research data platform.

This special issue presents an overview of Kashiwa Annex and its role in the mdx project, as well as other joint research with UTokyo.

AIDA, Kento

Professor, Information Systems Architecture Science Research Division, NII

Professor, School of Multidisciplinary Sciences, The Graduate University for Advanced Studies (SOKENDAI)

General Manager, Cyber Science Infrastructure Development Department (NII)

Interviewer

YAMADA, Tetsuro

Editorial Writer, The Yomiuri Shimbun

Interview

New Center for Open Science Launches

Kashiwa Annex to serve as a hub for managing and sharing research data

Every day, huge amounts of research data are generated and accumulated. If these "dormant" data are used effectively, it can lead to new knowledge. When combined with artificial intelligence (AI), new discoveries can be made. In March 2021, NII's Kashiwa Annex was opened to serve as a center for just this kind of research. Professor AIDA, Kento of NII helped establish Kashiwa Annex. We spoke to him about the significance of this new facility for promoting a more data-driven society.

Storage services to help in managing and sharing data

YAMADA What kind of facilities does Kashiwa Annex on The University of Tokyo's ("UTokyo") Kashiwa II Campus have?

AIDA NII is offering a range of both on- and off-site services to researchers. Kashiwa Annex is equipped with all the computers, networks, and other equipment necessary to provide these services.

A featured on-site service is the cloud-based research facility operated by NII. This is powered by approximately 100 high-capacity servers.

As a joint-use project service for off-site users, the facility also houses the storage for NII's Research Data Cloud (NII RDC) set up last fiscal year. (See note *2 on p. 7.) This platform enables researchers to manage their own data.

YAMADA What is the significance of this storage service?

AIDA NII was never involved in storage services before now. We provided interfaces and services for data management, but the data had

to be stored on the university's own storage or with a contracted cloud provider. So NII's research data platform, which connected to the university's data storage, could be used to manage the research data stored by the universities.

However, while large universities can afford to run their own storage or pay for cloud services, only a limited number of universities could do this successfully, due to a lack of funds or IT skills. Many small and medium-sized universities cannot have their own storage. To address this problem, NII set up a standard storage service for universities to keep their data on. This is a completely new service.

YAMADA So, there was quite a strong demand for a storage service?

AIDA Universities with their own supercomputers can do a lot by themselves, which is good, but fewer than 1,000 institutions in Japan are connected to SINET. Plus, only a few of these have the technology to operate a supercomputer. Most lack the equipment or personnel. So the idea is that NII, as an inter-university research institute, can now give them the support they need.

Managing and sharing research data for new research

YAMADA What kind of research can we expect them to do using the new storage service?

AIDA All kinds of research are possible, but the service is primarily designed to help with managing and sharing data after research is completed. For example, it helps to prevent research data disappearing from a PC left in a laboratory. And since research data often include sensitive information these days, the data needs to be securely managed by carefully restricting access. Our goal is to create a culture, or more precisely a system, for managing research data rigorously.

Next, after research data can be properly managed, the service will facilitate the use of accumulated data for new research, even by other researchers. I think this ability will expand the use of research data tremendously.

Therefore, the two main aims of Kashiwa Annex are data management and data utilization. One of NII's jobs is to create a common infrastructure for achieving these goals.

YAMADA What kind of new research do you envision will be done using the stored data?

AIDA I can give you an example of another project from my own experience, involving medical big data. NII's Research Center for Medical Bigdata (RCMB) collaborated with six Japanese academic societies in the medical field to build a cloud platform for performing AI-based analysis of images collected from hospitals across the country. In this project, the doctors who collected the data and the AI and image analysis experts who analyzed the data all used the same platform for their research.

About 30 joint research topics were launched on the platform, giving rise to a large interdisciplinary community. The process of collecting data brings together the people who provide and analyze the data, leading to the emergence of a variety of ideas and research challenges. This is a good example of how a common platform can generate interdisciplinary research. We hope that a similar process will happen in Kashiwa.

YAMADA So, if NII simply provides a stage for research, rather than telling people what to do, interesting research will happen naturally.

AIDA No, that's a little overly optimistic. It won't work if we just say, "Here are the facilities, now take care of the rest!" Unless NII makes a commitment, by entering into a joint research effort and thinking over the problem together, the creativity will not happen. Take the case of medical big data! Although we don't have any doctors at NII, we do have numerous experts in image processing and AI, so there was a keen desire to center the project on us. It is not enough to just provide the platform and leave the rest to others—we need to study other fields and make the effort to connect people on the common platform.

YAMADA So, to generate research findings, NII needs to build the stage, as well as perform on it together with people from other disciplines.

NII wants to serve as a hub that contributes to research matching

YAMADA Is there a system at Kashiwa Annex for NII to provide support and coordination?

AIDA That is something we need to create, I think. Kashiwa Annex only started operating this past spring, so we are not there yet. The fact is that the people who want to collect data are not computer experts. We need to create a mechanism for matching people who say, "I have this kind of data," with others who say, "I want to do this kind of analysis." We need to seek out such projects. Our approach to advancing research should be, "Look, here are some data. What about doing this kind of research?"

YAMADA You're saying that NII needs to approach researchers and uncover projects. Do you get any requests to collect totally new data using mobile devices, for example?

AIDA Yes, we get a variety. In addition to information about the movement of automobiles and people, in the field of agriculture, we

are collecting data on the movement of cattle over grazing land. And in the medical field, there is a move to collect and analyze human biometric information.

I think the common platform will spur a variety of new joint research efforts and collaborations. NII wants to serve as a hub for such initiatives. One of the strengths of Kashiwa Annex is that UTokyo's Information Technology Center (ITC) is located in the same building and on the same floor as NII. In fact, mdx, UTokyo's new "platform for the data-driven future" is based here. This mdx platform also serves as a hub, connecting universities and research institutes across Japan using the ultra-high-speed SINET network. Thus, we expect that a broad range of research will be spawned via Kashiwa.

YAMADA Are there any weak points or concerns that could become bottlenecks in achieving these goals?

AIDA As I mentioned earlier, simply installing equipment is not enough; there must be people to provide different kinds of support. However, since the number of people capable of interdisciplinary support and coordination is limited, how to cultivate and attract such people will be a challenge.

Very few people can handle a wide range of tasks, from building computer infrastructure to collaborating on research with experts in other fields. Academics tend to be rated on how many research papers they publish, but the work of creating and maintaining data infrastructure does not contribute to this. Therefore, I think it is important to properly appreciate the abilities of such people and provide them with suitable positions in academia.

The success of Kashiwa Annex will be measured not so much on how many people come and go from the building but rather on how the facility enables the effective utilization of research data. I would be happy to hear researchers complaining that there are not enough computing facilities available. That would mean that the facility was being used to maximum advantage.

Photography by MORI, Takahiro

A Word from the Interviewer

The fact that data utilization is increasingly influencing research capabilities is not confined to materials science or life science. Even in the humanities and social sciences, this trend will eventually give rise to unexpected investigation methods and research findings driven by big data. Japan is reported to lag other industrialized countries in the digitalization of administrative procedures and corporate activities. Fortunately, though, this does not apply in academia. I hope that NII's Kashiwa Annex will play a leading role in the promotion of data-driven research for a new era.

YAMADA, Tetsuro

Editorial Writer, The Yomiuri Shimbun

Joined the Yomiuri Shimbun in 1990 after graduating from The University of Tokyo. In 2006, he studied at the Massachusetts Institute of Technology (MIT) as a Knight Science Journalism research fellow. After working in the economics and science departments of the newspaper, and as a special correspondent at its Washington bureau, in 2018, he was appointed Science Department Chief. Since 2019, he has served as an editorial writer (on science and technology).

NII Kashiwa Annex (opened March 2021)

NII Kashiwa Annex facility information		
	Breakdown (reference)	
Equipment rooms	1F 480.80 m²	R&D Room (NW): 194.40 m² Research Server Room 1: 286.40 m²
	2F 669.60 m²	R&D Room (Cloud): 334.80 m² Research Server Room: 334.80 m²
Experiment laboratories	5 laboratories (1 on 2F, 4 on 3F, each of 64.80 m²)	
Joint research rooms	4 rooms on 3F (2 of 129.60 m² and 2 of 121.60 m²) * One of these rooms is divided into 16 cubicles with simple partitions to serve as a shared office.	
Meeting room (conference room)	1 room on 4F (49.36 m²)	
Other facility information	• Emergency generator (500 kVA) • Facility staff on duty 24 hours a day	

Prof. KAWARABAYASHI, Ken-ichi of NII setting up a PC in his laboratory at Kashiwa Annex

Interview

The Aims of mdx

Expectations of collaborative synergy between The University of Tokyo and NII Kashiwa Annex

TAURA, Kenjiro
Director, Information Technology Center, The University of Tokyo
Professor, Graduate School of Information Science and Technology, The University of Tokyo
Visiting Professor, NII

Interviewer
YAMADA, Tetsuro
Editorial Writer, The Yomiuri Shimbun

mdx is a new "platform for the data-driven future" centered at the Kashiwa II Campus of The University of Tokyo (UTokyo), jointly operated by nine universities, including UTokyo, and two research institutes, including NII. Its stated purpose is to facilitate the utilization of research data. To find out about the features and potential of mdx, we interviewed project leader TAURA, Kenjiro, director of UTokyo's Information Technology Center (ITC).

mdx is set to accelerate interdisciplinary big data research

YAMADA How did this academic data center come to be located at Kashiwa II Campus?

TAURA It all came about from the fact that former UTokyo president GONOKAMI, Makoto and NII director KITSUREGAWA, Masaru both felt a strong need for a new public base to serve our increasingly data-driven society. UTokyo also decided to relocate the ITC here, on the same site as the new mdx "platform for the data-driven future."

YAMADA The UTokyo ITC is already the key body of the Joint Usage/Research Center for Interdisciplinary Large-scale Information Infrastructures (JHPCN),*1 which links up eight universities.

TAURA The ITC's mission is not limited to UTokyo; it is a nationwide mission for setting up supercomputer infrastructure. Supercomputers have become widely used for computation by science and engineering researchers, as well as for informatics research, but in recent years there has been a growing demand to use such information infrastructure as a hub for interdisciplinary research. In response, we are providing, to the whole of Japan, a facility that no single university could possess by itself. At the same time, we are enabling researchers who are not experts in IT or data systems to use supercomputers. All this serves to promote joint research.

Up to now, supercomputers were mainly used for running simulations in science and engineering-related fields. The goal of the mdx platform is to extend the use of supercomputers into other fields, by focusing on data science and data utilization, thereby giving a wider range of users the opportunity to take advantage of supercomputing resources.

YAMADA Do you envision that mdx will help to promote more interdisciplinary studies and greater use of big data across all research fields?

TAURA Yes. Some researchers who have used supercomputers want to collect data generated from computations and experiment results to create databases, for later analysis and also to share them with other researchers. The aim of mdx is to enable and facilitate the kinds of things that were previously impossible or difficult with supercomputers.

Supporting academic research through collaboration with NII RDC (Research Data Cloud)

YAMADA Some researchers might not even know that such a data resource exists, or where to find it.

TAURA We need to work on different layers at various depths. It is not a problem that we can fix by creating a single platform like mdx. My view is that we need to create a system for cataloguing and searching through data on a nationwide basis. For example, data on medication dosage and medical practices, data

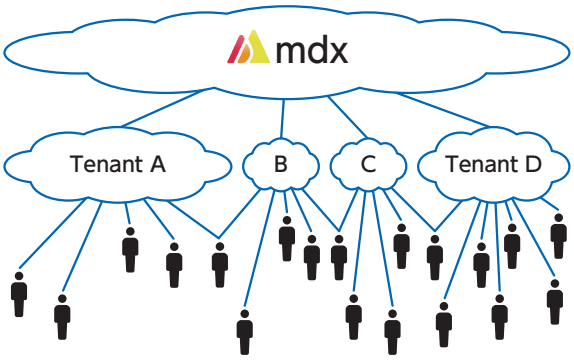


Fig. Virtual environments built on mdx and how they are used
The high-performance virtual environments offer separate, highly secure environments for different groups of users.

from IoT sensors and experiments, and data on weather and other observation data need to be collected and efficiently shared within each discipline and made available for easy processing and computation. We need a platform that can enable all of this.

YAMADA How will you work with NII?

TAURA NII's Research Data Cloud (NII RDC)*2 project has been in full operation since FY2021. One of the platforms running on RDC, GakuNin RDM, is used to manage and share research data among researchers. Researchers can also easily publish the data they manage by connecting to JAIRO Cloud,*3 a cloud service for institutional repositories.

One of the roles of mdx is to provide reliable storage for the data shared and published through the NII RDC. Another one is to provide the computing power needed to process any big data obtained through searches. The overall challenge for the NII RDC and our mdx platform is to provide a service that is easy to use for everyone. We want something suitable for all kinds of users, from highly skilled IT professionals to casual users who just want to do a quick data analysis in a web browser.

Connecting research institutes to establish a nationwide academic infrastructure

YAMADA What are the advantages of mdx from the viewpoint of computer and network technology?

TAURA Up to now, supercomputers were typically set up in one large room that all the users would access to use. Although there were certain measures to prevent other users from viewing your data without permission, the degree of separation between users was weak. For this reason, many users were hesitant to upload personal data, medical data, and other sensitive data, out of security concerns.

mdx works very differently, by allowing users to create a completely independent virtual computing environment. (See figure.) For applications in which security is vital, users can create a separate, highly secure environment and they can basically customize the environment however they like. This is the biggest difference compared to the past. It is quite an innovative approach to cloud computing in academia.

YAMADA Do you expect to see a lot of synergy between UTokyo and NII?

TAURA Given that NII's Kashiwa Annex and UTokyo's ITC are in the same building, we expect to have close personal exchanges. NII connects universities across Japan through the Science Information Network, SINET. And as I mentioned, researchers share their data through the NII RDC cloud. I believe that close cooperation between NII and ITCs will be indispensable in the process of creating a national integrated academic infrastructure that connects these networks with mdx and the supercomputer systems at specific universities. You can look forward to continued progress on this front.

Photography by MORI, Takahiro

[Glossary]

*1 JHPCN = Joint Usage/Research Center for Interdisciplinary Large-scale Information Infrastructures is a network for joint research operated by the supercomputer centers of eight Japanese universities: Hokkaido University, Tohoku University, The University of Tokyo (core institution), Tokyo Institute of Technology, Nagoya University, Kyoto University, Osaka University, and Kyushu University. JHPCN accepts applications for joint research projects from all over Japan and allocates computing resources and time (mainly supercomputers up to now) to selected projects.

*2 The NII RDC is an infrastructure service provided by NII to support the lifecycle of research data. It began full-scale operation in 2021. It consists of a research data management platform (GakuNin RDM), a publication platform (WEKO3), and a search platform (CiNii Research).

*3 JAIRO Cloud is a cloud service developed by NII based on WEKO. It is used to deliver repository services to research institutions. It was launched in FY2012. Since July 2016, the service has been operated jointly by JPCOAR and NII.

TAURA, Kenjiro

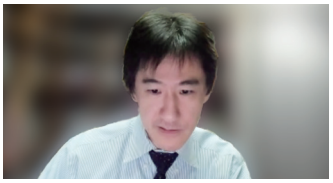
Earned his PhD from The University of Tokyo Graduate School of Science in 1997. He began working as a research assistant at the same university in 1996, progressing to lecturer in 2001, associate professor in 2002, and professor in 2015. Since 2018, he has served as Director of UTokyo's Information Technology Center. He is also currently a visiting professor at NII. His research specialties include parallel processing, high-performance computing, programming languages, and system software. He is pursuing research on balancing high-performance, large-scale parallel processing with ease of programming and use.

Roundtable

Building a New Foundation for Knowledge

Developing infrastructure for scientific and technological progress

YASUURA, Hiroto
Chief Cyber Science Infrastructure Director, NII;
Information Committee Chair, Council for Science and Technology, Ministry of Education, Culture, Sports, Science and Technology (MEXT)



TAURA, Kenjiro
Director, Information Technology Center, The University of Tokyo;
Professor, Graduate School of Information Science and Technology, The University of Tokyo;
Visiting Professor, NII

AIDA, Kento
Professor, Information Systems Architecture Science Research Division, NII;
Professor, School of Multidisciplinary Sciences, The Graduate University for Advanced Studies (SOKENDAI);
General Manager, Cyber Science Infrastructure Development Department (NII)

Since the services provided by Kashiwa Annex will serve the whole of Japan, collaboration and cooperation with universities and research institutes will be essential. Kashiwa Annex will also function to stimulate and accelerate research by connecting users to each other. The previously disparate ways of utilizing data will be reorganized and optimized to help achieve advances in science. We asked three of the key persons tackling this major challenge to speak about the current state and prospects of this initiative.

Supporting advances in data science

TAINAKA How did this new information infrastructure come about?

TAURA Up to now, large-scale computing and data infrastructure has mainly been used for running simulations in science and engineering fields. Now, however, we are seeing digital archiving in the humanities and other advances in IT and data utilization in the social sciences, medicine, pharmacy, education, and various other fields.

One of the factors driving data science development is the advent of artificial intelligence (AI). AI, especially deep learning, can be used to predict unknown outcomes based on data, even when the principles underlying a phenomenon are unknown. That is, even in the absence of equations, all kinds of things can be clarified if there are enough data. Another factor is the rise in the volume of data that can be acquired. Advances in sensor technology, for example, make it possible to obtain a wide range of observation data quite easily.

Up to recently, the Information Technology Center (ITC) and supercomputers around Japan have focused largely on conventional simulations, so there has been a mismatch in the way they were used. To address this growing user base, it was necessary to create a new kind of infrastructure, better suited to the new research methods. That is the root reason for the new developments.

YASUURA Observe a phenomenon, form a hypothesis about it using mathematical equations, and predict its behavior. Since Greek and Roman times, this is what science has been about. Most notably, the discipline of physics since the 18th century was developed through interaction between theoretical physics and experimental physics. Later, with the rise of computers in the mid-20th century, computational science emerged, making it possible to predict many kinds of phenomena by means of computation when numerical formulas are available. This is precisely how weather forecasting works.

Furthermore, since about end of the 20th century, the volume of available data has surged, with gigabytes (10^9 bytes) of data now the norm on PCs. When I was studying informatics in the 1970s, the biggest university computers only had around 1 to 2 megabytes (10^6 bytes) of memory capacity. Today, we are starting to see data capacities rated in exabytes (10^{18} bytes) and zettabytes (10^{21} bytes).

As we saw in the past with theoretical and experimental sciences, computational science, which involves humans formulating equations to perform simulations, and data science involving technologies like AI are influencing each other as they advance. Kashiwa Annex and mdx, as a "platform for the data-driven future," are working to create a mechanism to support data science.

AIDA Up to now, researchers have largely managed their own data, but this is no longer possible due to

the massive increase in data volume. This initiative was driven by the need for a common data management platform. To share data, it is essential to properly manage access privileges and control the scope of data sharing. However, setting up infrastructure for this purpose in a way that suits all the different research groups is difficult. The location of data was left to the judgment of users. NII's research data platform was designed to facilitate research data utilization by providing not just a mechanism for managing research data but also data storage.

TAURA It is difficult for researchers to collect the data they need for research by themselves, so there is a lot of value in sharing data. On the other hand, disclosing personal information or other highly sensitive data, or data related to the commercial development of materials or pharmaceuticals, is certainly difficult. I think it is vital to both enhance security and properly demonstrate the value of utilizing research data.

TAINAKA Is a national strategy necessary?

YASUURA As part of IT-related science and technology policy, in the 1970s, large computer centers started to be built at major universities. These cost billions of yen at the time, despite being less powerful than the PCs of today. Then, in the late 1990s, a supercomputer known as the "Earth Simulator" was developed to perform high-performance computations to simulate the atmospheric conditions of the whole planet. Later, a national project was launched to build an even faster supercomputer, leading to the "K" and "FUGAKU" supercomputers. In the field of computational science, it is very important for a country to have the technology to perform computations on par with the world's most advanced computing systems, as well as the personnel to operate the technology.

We now need to take this same approach in the field of data science. One challenge is how to create a system that is as good as a supercomputer in terms of stably maintaining extremely large volumes of data. However, it is not just a question of data quantity, but also how the data are used. As a country, we need technology for handling data in ways that match the characteristics of different types of data, for example, regarding the level of security, or whether the data are publicly disclosed. With the introduction of mdx, we are now starting to explore and think about the ideal academic data platform.

The appeal of data science lies in the possibility of making new discoveries by combining data from different disciplines. When different communities open their data to each other and study the data together, new discoveries can be made. Another challenge is creating a place to try out these possibilities.

Platform design and operation

TAINAKA What are the key points for designing a data platform?

TAURA How to connect people with each other is important. Data sets are all different in terms of quantity, and their numerical values have completely different meanings, so only experts in the field can judge the data's reliability and interpret the data. I feel that it is essential for the people in a discipline, who are knowledgeable about the data, to work on and discuss problems with people like us who know about computational methods and computers. Problems need to be addressed in a very fine-grained way like this.

AIDA How data are collected is also important. Up to now, users typically just loaded their data into a supercomputer to get their computation done. Since it is now generally assumed that data are directly transmitted from sensors, supercomputers need to have not only resources for calculation but also a way of receiving data. The format of the received data will vary widely, so different software will be needed for different projects. It is also necessary to think about variations for specific situations, such as when sensitive information needs to be sent in anonymized form.

Having a fast network and servers is not enough. A way of running such software securely is required, and this too must be created through discussions with users.

TAINAKA What about operation?

TAURA mdx is jointly operated by 11 institutions around Japan as a single common platform. We want to maximize the research findings obtained through mdx by taking advantage of the strengths and expertise of all the participating institutions, particularly by connecting research groups in various disciplines. (See the figure.) We are also working to create a new kind of service that enables any new machine purchased by one of the institutions to be shared by all the others. We want to expand this data science platform by producing more and more research findings.

TAINAKA mdx was officially opened only recently, in September 2021. How is it going so far?

TAURA Users can apply to use mdx from the mdx website. mdx is also accessible via the GakuNin authentication platform operated by NII, which links universities across Japan. However, many universities are not yet set up to use mdx, so we need to let them know about this new resource.

AIDA For security reasons, the services available via GakuNin can only be used at a university after the necessary permissions have been set up. Previously, GakuNin was

used mainly for resources that can be accessed at all universities, such as electronic journals. Since this is probably the first time that a service delivered through a computing platform like mdx will be used, we need to educate universities about the service.

TAURA We are also inviting general users to participate in mdx, so research laboratories and individuals are gradually starting to use the system. mdx is already being used in a wide range of fields, including AI, image processing, spatial information for smart cities, genomics, materials, and human flow data for economic analysis.

YASUURA We are still at the trial stage, dealing with a variety of issues. Nonetheless, researchers without access to powerful computers or data infrastructure can now easily apply to use mdx for their research. I would say that if everything goes smoothly for the next five years or so, we could call this a successful support initiative.

Making it work as a national infrastructure

TAINAKA How will mdx work alongside commercial cloud services?

AIDA Each community has its own ideas about where to store data, but I think it would be good if we could decide on guidelines for different data uses and implement them successfully.

TAURA Commercial cloud services are pricey and most vendors are based overseas, so everyone is using them out of necessity. That is the truth. Due to a lack of alternatives, they are using commercial cloud services for whatever they cannot handle with supercomputers. To what extent can we develop a national infrastructure to meet this need? The vital thing is to create an environment that is easy to use.

YASUURA Research is competitive. The first person to make a new discovery gets the glory, so it is essential to be No. 1. We need to create a solid foundation for the data field so that inadequate government support does not prevent researchers from producing findings.

The Information Committee of the government's Council for Science and Technology (MEXT), which I am currently chairing, is debating the question of how to create and maintain an academic information infrastructure for all fields of academia. In addition to FUGAKU and the Science Information Network (SINET), the committee will also be deliberating on mdx-based data infrastructure services.

Since the technology in this field tends to become outdated after six or seven years, we need to keep preparing budget requests while keeping in mind the future. It is particularly vital to make sure that our service delivery method is sustainable. In other words, we need to secure stable funding for personnel, electric power, and various other expenses. If mdx proves to be successful, we can even consider increasing the funding by an order of magnitude, for example to distribute the infrastructure across multiple locations. But for this, we need to demonstrate success.

The role of NII

TAINAKA What do you expect from Kashiwa Annex?

TAURA Well, we have many expectations for NII (laughs). Academic information infrastructure is currently scattered across various universities, so users need to have accounts with multiple universities. I would like to see NII build a data infrastructure platform that brings together all these physically dispersed services so that users can be presented with a single system.

For us, the most important thing is collaborating with Prof. AIDA and everyone else at Kashiwa Annex. We want to hold regular strategy meetings to discuss familiar topics like mdx, SINET, and the NII Research Data Cloud. We hope to build a close relationship with NII so that whenever a difficult problem arises, we can tackle it together by securing funding and launching a joint research project.

AIDA I would welcome that. When data are generated and analyzed over the network, the results can be valuable to society. Our goal with Kashiwa Annex is to create a single, streamlined resource to enable this. It is also very significant that the infrastructure to achieve this has been established in Kashiwa.

SINET now includes a mobile network, as well as a wired network, so we are also developing a way to automatically collect data via this mobile network and feed them reliably into the computing system. We want to combine these technological mechanisms like puzzle pieces and hook them up to the computational power of mdx.

Coincidentally, in our student days, Prof. TAURA and I were active in the same community. In the past few years, we have both worked on building and operating infrastructure, so I very much look forward to working with him on creating new infrastructure for Japan.

YASUURA Sharing research resources is vitally important, but since each community and university has its own unique culture, the problem is not so simple. As NII has worked on SINET and other projects, it has tried to overcome these limitations. Now, we are striving to implement a data version of this in Kashiwa. I think we are well positioned to play an important role in the operation of the new system.

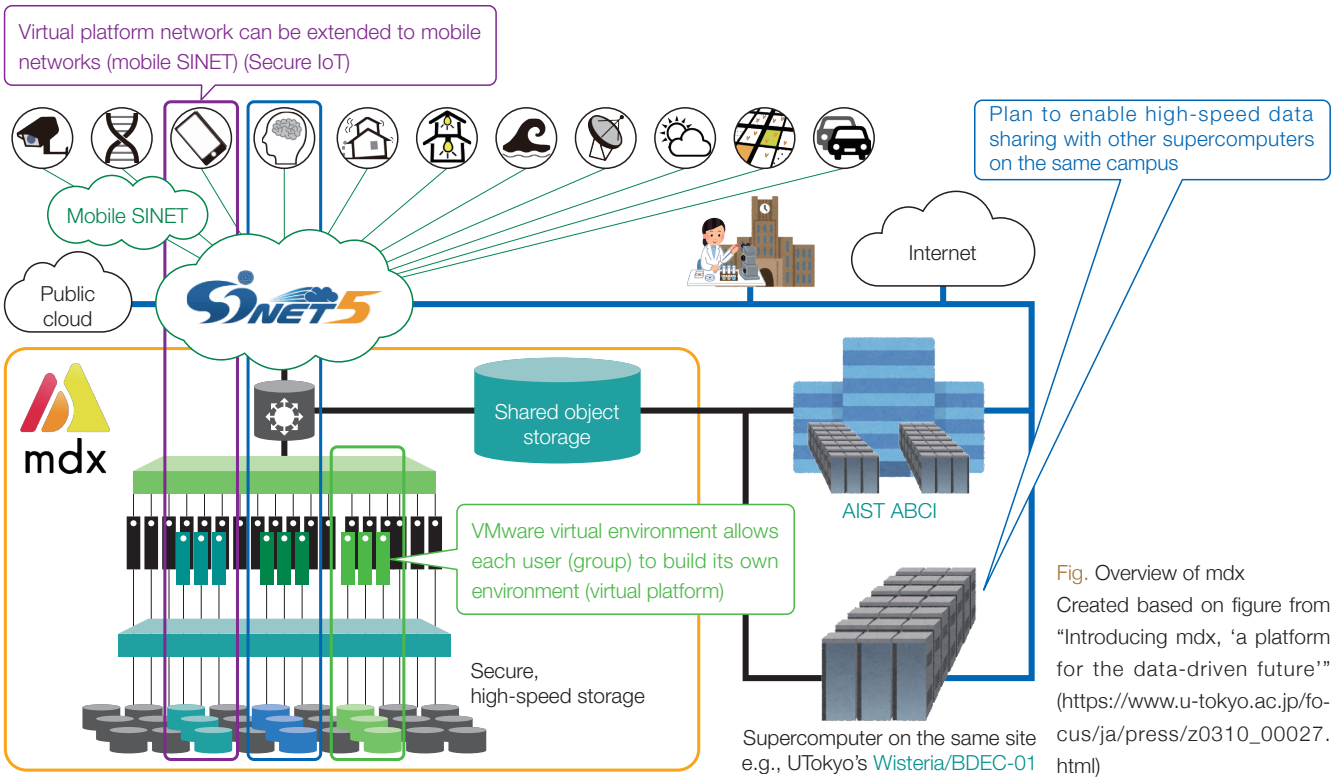


Fig. Overview of mdx
Created based on figure from "Introducing mdx, 'a platform for the data-driven future'" (https://www.u-tokyo.ac.jp/focus/ja/press/z0310_00027.html)

TAINAKA What about personnel development?

YASUURA This is a huge issue. The importance of this work must be better understood. In academic papers, it is increasingly common to find that the people who create and organize data are given as much importance as the authors. Research findings are produced through the joint effort of the people conducting the research and the people who handle the vital infrastructure that supports the research. It is essential that we foster this culture.

NII needs people capable of steadily maintaining and improving its cutting-edge services. Even at universities, there are many outstanding individuals who are suited to practical work more than research. We need to create an environment in which such people are considered as important as researchers, reflecting this even in salaries. Otherwise, science and technology in Japan cannot thrive.

TAURA As a researcher in the field of IT, it is frequently difficult to strike a balance between writing papers and helping people in other fields to produce research findings. Due to my knowledge of IT, I am often called on to help researchers in other fields. I want to do what I can to change the academic environment so that such key contributions are no longer overlooked.

Many classes in recent times have been held online due to the COVID-19 crisis. It so happened that students with IT expertise spontaneously came together to set up a system to help teachers and students in need. At UTokyo, this system has now taken root. If enthusiastic students who can sympathize with other users in this way are exposed to this kind of support work at the right time and in the right way, we can probably nurture the young personnel we need.

AIDA Whenever data are collected, the various people involved with the data will also come together, so connections are naturally formed. This activity should help to expand such connections even further. As an infrastructure provider, I want to have many conversations with people in practical fields of research so that we work together to help solve each other's problems.

YASUURA Since the development of data science infrastructure directly influences scientific and technological accomplishments, it has become a focus of global competition. Although mdx is only our first step in this process, it is no exaggeration to say that Japan's scientific and technological progress over the next decade or more rests on its shoulders. We can also note that this initiative is more advanced than comparable efforts in other countries. I therefore want to make this project widely known to encourage as many researchers as possible to make use of it.

Interview by TAINAKA, Madoka;
Report by HIRATSUKA, Yuko

Joint Research

DIAS, an Earth Observation Data Platform

Research from Kashiwa Annex that is helping to solve social issues

IKOMA, Eiji
Project Associate Professor, Earth Observation Data Integration & Fusion Research Initiative, The University of Tokyo

KITAMOTO, Asanobu
National Institute of Informatics
Professor, Digital Content and Media Sciences Research Division, NII
Professor, School of Multidisciplinary Sciences, The Graduate University for Advanced Studies (SOKENDAI)

An important purpose of NII's Kashiwa Annex is to bolster research collaboration with The University of Tokyo (UTokyo). In this context, we present the Data Integration and Analysis System (DIAS), a national project rapidly recognized as increasingly important in the light of climate change in recent years. To learn about DIAS, we interviewed IKOMA, Eiji, a project associate professor at UTokyo's Earth Observation Data Integration & Fusion Research Initiative, who is leading the project from a systems perspective, and NII professor KITAMOTO, Asanobu, who is working to advance research from the viewpoint of Open Science.

DIAS is a platform for global environmental research

How can science find clues to help tackle global warming? DIAS is a big data platform to support researchers working on problems like this. It can store a total of around 50 petabytes (petabyte = 10^{15} bytes) of global environmental data. It also links to middleware for integrating and managing these data, to common APIs and application development platforms for utilizing various kinds of data, and even to services that researchers can use directly. The key feature of DIAS is its ability to support research activities through comprehensive data management in a way that facilitates predictive analyses and simulations.

Phase I of DIAS began in 2006.*1 The current Phase IV "Global Environmental Data Integration and Analysis Platform Project" (2021–) is being implemented mainly by the Japan Agency for Marine-Earth Science and Technology (JAMSTEC), with UTokyo and NII participating in support roles. Yet, DIAS did not start suddenly in 2006. Its origins go back some 40 years to the archiving of earth observation satellite data by UTokyo's Institute of Industrial Science (IIS). The DIAS of today is the product of all the activities that IIS has steadily carried out on its own over the years, from data collection to storage, involving handmade antennas for receiving satellite data and custom storage systems (tape archives).

As Professor IKOMA explains, with the aim of "creating public value through data," the IIS has worked for the past 40 years on receiving various kinds of satellite data for earth observation, building its own system for storing the collected data, and comprehensively managing the data for use in research. "Fundamentally, what we are doing now has not changed," he says.

Kashiwa Annex is ideal for handling huge volumes of data

The DIAS system is currently spread out across four locations: the ISS in Komaba, Tokyo; UTokyo's Kashiwa II Campus; Hokkaido University; and the Kitami Institute of Technology. (The Kashiwa II Campus is a home to both UTokyo and NII's Kashiwa Annex.) Communications between these sites are carried over NII's Science Information Network (SINET). External data too, from sources such as the Himawari meteorological satellite and XRAIN (a real-time precipitation observation system operated by the Ministry of Land, Infrastructure, Transport and Tourism) are transferred via SINET and stored in DIAS.

In describing the importance of collaboration with Kashiwa Annex, Prof. IKOMA says, "Aside from simulation data, observation data from Himawari and elsewhere flow into DIAS in real time every day, at the rate of terabytes (terabyte = 10^{12} bytes). For stable operation, the power of SINET for transmitting such huge volumes of data at high speed is vital. It would be difficult for the university to operate such a massive data system on its own. Thus, using the Kashiwa II Campus, with its direct connection to SINET, is indispensable for DIAS."

Professor KITAMOTO, who has been involved in DIAS since the earliest days of the project, stresses, "Kashiwa Annex has space for large-scale equipment and it is well equipped with power facilities. It ensures sufficient availability to support such a large-scale system."

As of August 2021, there were 9,416

registered DIAS users, about five times the number in FY2015. The cutting-edge facilities at NII's Kashiwa Annex support the backbone of DIAS, which has seen dramatic growth over the past few years, both in data volume and numbers of users and accesses.

Open Science, personnel, and other future issues

The opening of Kashiwa Annex has meant expanded hardware support for DIAS. There will also be a stronger focus on Open Science in the coming years. Professor KITAMOTO explains, "With global environmental data, there are relatively few privacy issues. Much of the data, like environmental and disaster-prevention data, are highly public, so researchers feel quite free about sharing data among themselves. This makes it easier to engage in "open science." Until now, however, data licenses were not always clear, a factor that hindered openness at times. Currently we are making improvements, such as attaching a Creative Commons (CC)*2 license to data sets."

According to Prof. KITAMOTO, the move to open data among researchers has rapidly grown stronger in recent years. In response to this trend, in 2017 DIAS began assigning a unique identifier, known as a Digital Object Identifier (DOI), to data sets. One of the challenges with open data is that it is difficult for data users (e.g., paper authors) to see the contributions of the people who created the data. Assigning DOIs makes it easier to see how the data are really used, as well as how the data contribute to research.

At the same time, Prof. IKOMA says that the future activities of DIAS will focus on measures to combat climate change issues. One of the most urgent requests of local governments is prediction of river flooding due to typhoons and localized torrential rains. For example, in the Kuma River area of Kumamoto Prefecture, a project has begun to determine the level of river flooding risk in real time, using DIAS to analyze images collected from live cameras installed in about 2,700 locations across Japan by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT). The images from the live cameras can be arranged in order of risk level or images can be checked in real time to identify areas of rapidly increasing risk. (See figure.)

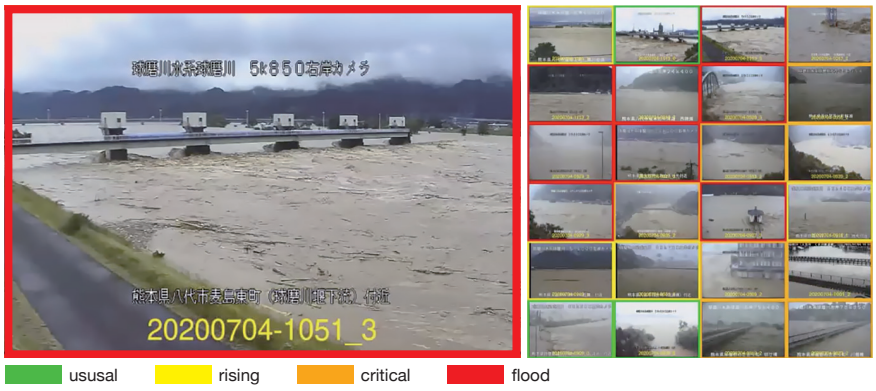


Fig.: Automatic recognition of flooding images based on deep learning (Kuma River water system). Live camera images are analyzed in real time using deep learning, and flood conditions are displayed in different colors.

This system is expected to be further expanded, by overlaying water level observation data and weather forecast data on DIAS, to enable the forecast of imminent river flooding and to check the state of evacuation sites through links with local government systems. There are also plans to apply this system in various other ways, for example, for flood prediction, disaster information portals, and linking climate data with tourist services.

However, a shortage of personnel presents a problem. It is undeniably difficult to secure suitable people because project personnel are required to serve a wide range of functions, from setting up systems and managing applications and services that run on the system, to supporting the researchers who use the system. On the bright side, the level of interest in climate change and sustainability among the younger generation is clearly higher than it was a few years ago.

"We need to show that involvement in DIAS will lead to a solid career path," says Prof. IKOMA. Our collaboration with Kashiwa Annex is an important step toward making DIAS a more appealing data platform.

(Reporting and writing by GOMI, Akiko)

[Glossary]

*1 Started as "Data Integration and Analysis System," a research project commissioned by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) (2006-2010), then continued with Phase II "Global Environment Information Integration Program" (2011-2015) and Phase III "Program to Promote the Development of Earth Environmental Information Platform" (2016-2020).

*2 CC: International rule for distribution (reuse) of copyrighted works in a way that reflects the author's intentions.

IKOMA, Eiji

Earned his PhD in information and communication engineering from The University of Tokyo School of Engineering in 2000. Having served as assistant professor at the university, he currently works as a project associate professor at UTokyo's Earth Observation Data Integration and Fusion Research Initiative. Since 2016, he has also served as a Senior Scientific Research Specialist at the Research and Development Bureau of the Ministry of Education, Culture, Sports, Science and Technology (MEXT). His fields of specialization are information engineering and database engineering.

NII NEWS TOPICS

News & Topics　Period: September 1 (Wed.) to November 24(Wed.), 2021

News Releases

■November 24, 2021
Discovery of the indirect network with the smallest theoretical diameter in “Graph Golf,” a competition to find graphs leading to the efficient design of supercomputers: Effectiveness in configurations where network switches are used for indirect connections between CPUs

■November 17, 2021
Japan Data Catalog for the Humanities and Social Sciences (JDCat) goes into full operation enabling data cross-searching in the humanities and social sciences

■November 15, 2021
Development of technology to detect critical problems in autonomous driving using simulation: Efficient automatic searching of potential problems by exploring whether problems will occur

■November 4, 2021
Solving optimization problems in control systems with deep neural networks : Building a new approach to discrete-time stochastic optimal control problems in constrained nonlinear systems

■October 20, 2021
Developing a technique for automatically and efficiently discovering a reliable gas turbine system design: Focus on logical specifications in black box optimization, applied to the design process for real products of companies

■September 22, 2021
Development of the SYNTHETIQ synthetic videodetector, a program for automatically determining fake facial videos generated by AI: Promoting research on AI video generation, fake media detection, and media reliability assurance

Awards

■November 3, 2021
A paper by Prof. JI, Yusheng (Information Systems Architecture Science Research Division) et al. wins 2021 IEEE Internet of Things Journal Best Paper Award.

■November 1, 2021
A paper by Assoc. Prof. MIZUNO, Takayuki (Information and Society Research Division) et al. wins 2021 Excellence Award at the 35th Annual Conference of the Japanese Society for Artificial Intelligence.

■November 1, 2021
A paper by GHALAMKARI, Kazu (Sugiyama Lab, Department of Informatics, SOKENDAI) and Assoc. Prof. SUGIYAMA, Mahito (Principles of Informatics Research Division) wins a 2021 Excellence Award at the 35th Annual Conference of the Japanese Society for Artificial Intelligence.

■November 1, 2021
A paper by Project Assist. Prof. LIU, Jia (Center for Strategic Cyber Resilience Research and Development) et al. wins Best Paper Award at the 2021 International Conference on Networking and Network Applications (NaNA 2021).

■October 27, 2021
A paper by Prof. INOUE, Katsumi (Principles of Informatics Research Division) and NII Visiting Prof. MORGAN, Magnin wins Best Paper Award at ILP2020-21 international conference.

■October 12, 2021
A paper by Project Assoc. Prof. ARCAINI, Paolo (Information Systems Architecture Science Research Division) et al. wins a Distinguished Paper Award at the 13th Symposium on Search-Based Software Engineering.

■September 17, 2021
Assoc. Prof. KANEKO, Megumi (Information Systems Architecture Science Research Division) wins the 2021 KDDI Foundation’s Contribution Award.

■September 9, 2021
A paper by Assoc. Prof. INAMURA, Tetsuya (Principles of Informatics Research Division) et al. wins the 9th Advanced Robotics Best Survey Paper Award from the Robotics Society of Japan.

NII Information

■November 15, 2021
NII Global Science Campus (GSC), 2022 “Master of Informatics” Program; applications open.

■October 27, 2021
“DX Symposium for Educational Institutions,” archive site opens.

■October 21, 2021
Latest book in the NII Series “Big Data Pioneers Medical AI” is published (by Maruzen Library).

■October 20, 2021
Menu of GakuNin cloud services expands; user application and services menu are now more convenient.

■October 6, 2021
Registration opens for “Frontiers of Informatics,” NII’s 2021 public lecture.

■October 5, 2021
PR magazine “NII Today” No. 92 “From Personal Information to Privacy” is published.

■October 1, 2021
NII starts accepting applications for FY2022 open joint research programs.

■September 1, 2021
“Films of the Great Kanto Earthquake of 1923” website is launched.

Events www.nii.ac.jp/event/2021

Events www.nii.ac.jp/event/2021

■Cyber Symposium on Online Education and Digital Transformation (DX) at Universities and Other Institutions, "DX Symposium for Educational Institutions"

- November 19, 2021: Online Symposium No. 43
- October 29, 2021: Online Symposium No. 42
- October 8, 2021: Online Symposium No. 41
- September 17, 2021: Online Symposium No. 40
- September 3, 2021: Online Symposium No. 39

■ November 22, 2021

NII-IDR User Forum 2021 (online)

■November 17, 2021

NII Service Briefing: "A Preview of Our Next-generation Academic Research Platform! What is Changing?"

■November 1, 2021

23rd Library Fair & Forum (booth exhibit & forum)

■November 1, 2021

University guidance session for Department of Informatics, School of Multidisciplinary Sciences, The Graduate University for Advanced Studies

■October 24, 2021

Inter-University Research Institute Corporation Symposium 2021: Space, Material, Energy, Life, Information, Human Culture; "Let's Enjoy Knowledge from the Frontlines" (online)

■2021 Public Lectures, "Frontiers of Informatics"

- Viewable from October 10, 2021
  Lecture 1: "Searching Data for Hidden Regularities!—Structures of Woven Patterns and AI" by SUGIYAMA, Mahito
- Viewable from November 1, 2021
  Lecture 2: How to Assess the Reliability of AI Products?—The Wonders of AI and the Difficulties of AI Products" by ISHIKAWA, Fuyuki
- More lectures will be posted successively from December 2021.

We want your feedback!

To help us improve the magazine further, please tell us what you think about the magazine, using the URL below or the QR code on the left. We look forward to hearing from you.

www.nii.ac.jp/today/iken

Bit (NII Character)

CiNii Research is now fully operational.

CiNii Research, a service that enables users to search through research data in articles, books, and other publications through a unified interface, was released on April 1, 2021.***1**

A preview version of CiNii Research was released in November 2020, but we have now added advanced search, search using "OR", "NOT", and other operations, as well as RDF and other API functions, to deliver a more comprehensive search experience.

As of October 4, 2021, the functionality of CiNii has also been enhanced by with the inclusion of OpenSearch, JSON-LD, and ResourceSync as APIs, as well as an export function for integration with literature management software.***2**

As of the end of September 30, 2021, there were approximately 320,000 research data items available on CiNii Research. Although the number is still small compared to research papers, the number has grown steadily since the service was first released, when there were approximately 180,000 data items. To ensure the promotion of Open Science, we are working to continue increasing the number of databases and the quantity of research data that are searchable.

In the spring of 2022, we also plan to integrate CiNii Articles, to further increase the power of article searches, improve functionality, and implement more advanced internal information.

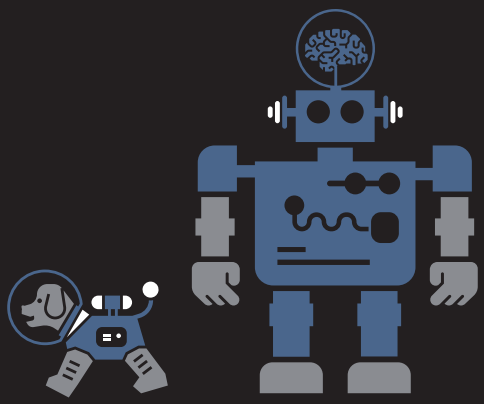
If you have been using the CiNii Articles API or linking to the CiNii Articles site, please switch to using the CiNii Research API or update the link to connect to the CiNii Research site.

***1** CiNii Research (https://cir.nii.ac.jp/)
***2** Integration of CiNii Articles into CiNii Research (https://support.nii.ac.jp/ja/news/cinii/20210706)

Essay

Dialogue in the Dark

INAMURA, Tetsunari
Associate Professor, Principles of Informatics Research Division, NII
Associate Professor, School of Multidisciplinary Sciences,
The Graduate University for Advanced Studies (SOKENDAI)



Have you heard of an event called "Dialogue in the Dark"? It offers an opportunity for able-bodied people to experience how people with visual impairments experience everyday life. You walk in a park or shop at a store that is totally dark, with no visible lights at all. When you are thrown into such a situation, your sense of hearing and touch get sharper (especially touch), and you feel as if the relationship between sensation and action is being reconstructed inside the brain at high speed.

Apart from heightened senses, you realize another thing: how important the presence of other people is. Since individual able-bodied people cannot do anything on their own, they form a group to move through the darkness, led by a person with visual impairment. The focus of your hearing, touch, and other senses is not on physical objects or the environment, but rather on the presence of others. The person moving ahead calls out to those behind, so that they do not fall into the stream that flows through the park. In a store, not knowing what is displayed, you ask a person with visual impairment playing the role of a shop assistant to hand over items. Naturally, you need to sharpen your sensitivity to surrounding objects to avoid bumping into walls, but most of your attention is on the presence of others. Only in the darkness does it become clear why the event name features the word "dialogue."

Humans have such a high degree of adaptability that they can cope with dramatic changes in conditions. But how would a robot deal with a similar situation? As I walked in the dark, I suddenly switched to "researcher mode." If its camera were damaged, perhaps the robot would begin to reconstruct its model of the world using other sensors. Using the model of the world it learned when its camera was working, the robot might be able to re-create visual images based on distance sensors. Although we can envision robots moving around in the dark nimbly, we cannot imagine them starting to actively interact with each other. That is because they can get by without dialogue.

During the dark days of the COVID-19 pandemic, humanity lost the modality of face-to-face communication. We could hear the voices of other people and know they were present, but the icons of their faces prevented us from physically sensing their presence. Although we never shook hands or hugged each other very often, we still felt a palpable loss of bodily presence. There was no alternative sensory modality to counteract this "darkness," and despite my best creative efforts, I failed to come up with a decisive solution. However, the thing that may extend a helping hand in the future to resolve this difficulty could be a robot, which ironically does not require the same kind of interaction that we do.

In recent years, we have witnessed remarkable improvements in the natural language processing capabilities of AI. However, a variety of challenges still stand in the way of linking body-based sensory modalities and experiences with language processing systems to achieve dialogue. One of these is the presence and necessity of "others." Even today, as I converse with others online and with our webcams on, I keep imagining a gentle robot that needs others, that can become a strong ally of humans, and that can illuminate the "darkness" of that sensory void.

Weaving Information into Knowledge
NII

National Institute of Informatics News [NII Today] No. 93 Dec. 2021
Published by National Institute of Informatics, Research Organization of Information and Systems
Address: National Center of Sciences,2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430
Contact: Publicity Team, Planning Division, General Affairs Department
TEL: +81-3-4212-2028 FAX: +81-3-4212-2150
E-mail: kouhou@nii.ac.jp

Publisher: KITSUREGAWA, Masaru
Editorial Committee Chair: KAWARABAYASHI, Ken-ichi
No. 93 Editorial Supervisor & Editorial Committee Member: TAKEFUSA, Atsuko; KOMIYAMA, Yusuke
Editorial Committee Member: IKEHATA, Satoshi; KANEKO, Megumi; MIZUNO, Takayuki (in Japanese syllabary order)
Production: TAINAKA, Madoka; IKEDA, Akiko/ KATAYOSE, Masashi (Sci-Tech Communications Inc.)
Design and DTP: KURAHASHI, Hiro (MATZDA OFFICE CO. LTD.)
Cover illustration: ICHIMURA, Joe