National Institute of Informatics News ISSN 1883-1974 (Print) ISSN 1884-0787 (Online)

NII Today

57
Mar. 2016
National Institute of Informatics News

NII Interview

Toward an Era of Open Science

The Forefront of Open Data

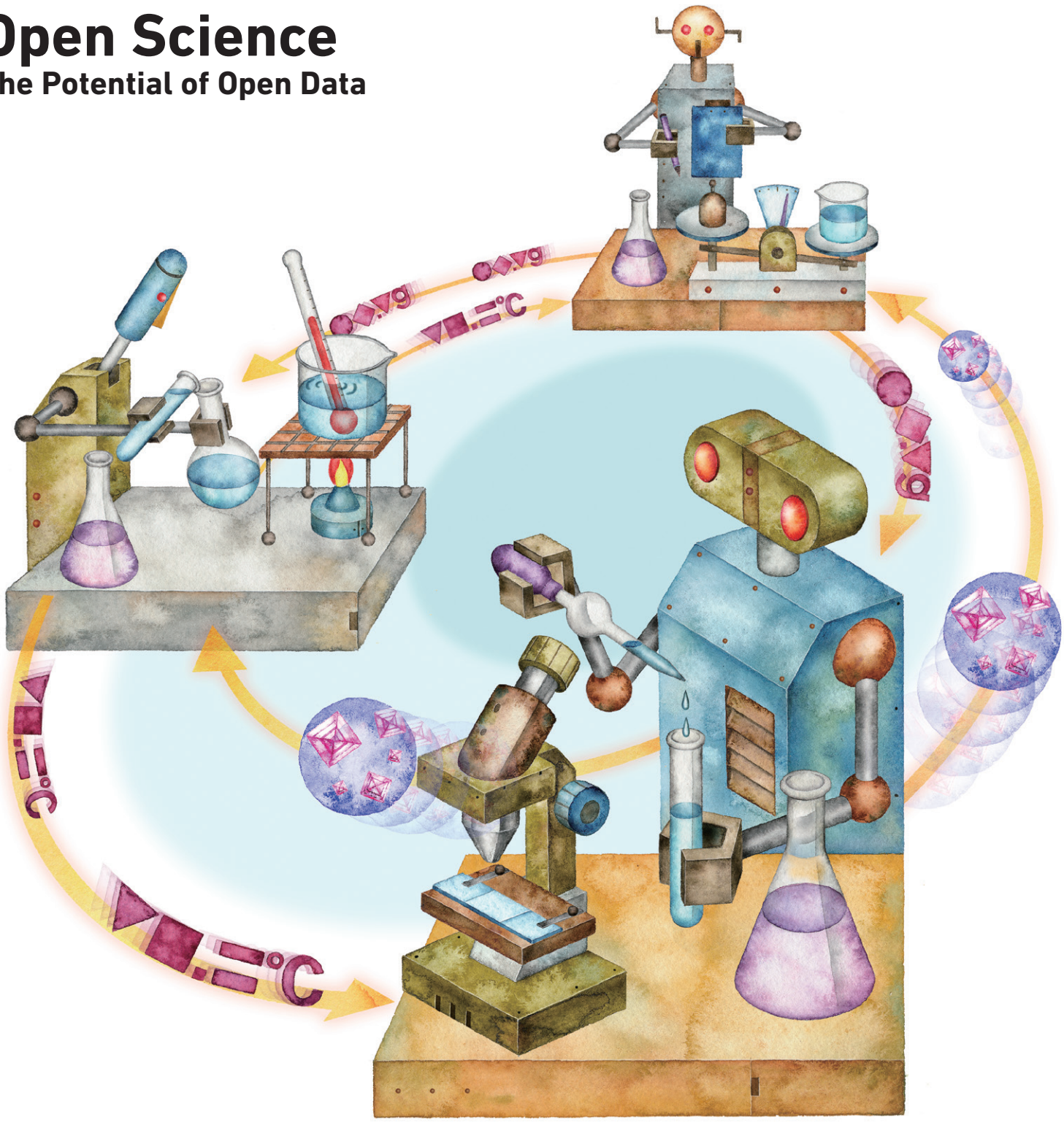
From Institutional Repositories to Open Science

Making Use of Real-World Data in Informatics Research

Feature

Open Science

The Potential of Open Data



This English language edition
NII Today corresponds to
No. 71 of the Japanese edition

Toward an Era of Open Science

What is NII's role in data sharing?

Masaru Kitsuregawa (Director General, National Institute of Informatics)
Interviewer: Junichi Taki (Editorial Writer and Senior Staff Writer, Economic Commentary Department, Nikkei Inc.)

Moves have begun to release and share the experimental data on which scientific literature is based. For the past several hundred years, information has been shared through academic journals and conference presentations, but this is changing dramatically due to the development of information and communications technology. This move toward "open science" has the potential not only to accelerate research but also to act as a catalyst for producing new knowledge by bringing together the knowledge produced by diverse researchers in a myriad of disciplines. The National Institute of Informatics provides IT infrastructure that supports Japan's scientific research, and I asked Director General Masaru Kitsuregawa about the role NII intends to play in open science.

Masaru Kitsuregawa

Taki: Open science is a phrase we hear a lot these days.

Kitsuregawa: There are two issues in open science: open access journals and open research data. Until now, readers have paid money to subscribe to journals containing scientific literature. In the case of open access journals, which are spreading worldwide, the authors pay money to the publishing company and the articles can be read by the general public free of charge.

On the other hand, open research data are the goal of a movement to release the data on which articles are based along with the articles. If one has the data, it becomes easy to reproduce the claims made in an article. As a result, many researchers are soon able to make use of an article's conclusions and data, and this can accelerate scientific progress and innovation. Open research data can also be expected to have the effect of reducing submission of articles that are not reproducible. If data are available, subsequent people will make use of the data in their research, which will not only speed up research but also avoid redundant investment and thereby promote efficiency.

The debate about open access journals appears to be drawing to a close for the time being, and so open research data have become the hot topic.

Taki: Making data publicly available is important, but there does not appear to be any incentive for researchers to do so.

Kitsuregawa: Digital object identifiers (DOIs) are numbers used to identify articles, and the practice of also assigning identification numbers to data and citing the data used when writing an article has already started. This trend respects the researchers who have produced valuable data and creates an incentive to publish data.

However, data are more difficult to evaluate than articles. Standards for data accu-

racy vary depending on the purpose of use, and the correctness of data may also vary if the way in which they are used changes. Perhaps data cannot be evaluated in the same way as articles.

I understand the feelings of researchers who want to retain possession of the data that they have worked so hard to create, but scientists are obliged to guarantee the reproducibility of their articles. The premise is that you will publish what other scientists publish. The question of how to achieve a sense of fairness could be broadly considered an issue of diplomacy. There is an organization called the Research Data Alliance (RDA) established by research institutions from countries including Japan, the US, and the EU that is holding discussions about the kind of values that can be newly created by data sharing.

Taki: How will NII respond to the trend for opening up data?

Kitsuregawa: NII supports places such as university libraries in operating "institutional repositories" for the collection, storage, and use of scholarly information of universities and other research institutions. NII offers a shared repository service called JAIRO Cloud, which is used by 465 universities and research institutions in Japan. I think that universities would be delighted if this service were expanded so that data could also be stored.

The nature of data differs between fields such as astronomy, high-energy physics, genome analysis, and materials science, and data handling practices also vary. Storing data presents different difficulties than storing articles, and issues such as what to do with the metadata that are assigned to indicate data content, for example, will have to be resolved through consultations with researchers in each field. It might take some time, but we have to move forward one step at a time.

Taki: So NII will support research by providing storage services.

Kitsuregawa: That's right. But with data come the programs used to analyze the data. These programs also have to be stored in order to guarantee the reproducibility of analyses, and this is also quite taxing in terms of technology. As everyone will have experienced with their personal computer, when an OS version changes, application software often stops working.

Given this, in the long term, NII will work together with researchers in each field to examine ways of storing data and to investigate and build a research platform containing the programs that handle the data.

This might be regarded as the world of what is called Science 2.0.

The general trend is for scientific research to be placed on IT platforms. Why does everyone use Amazon's cloud service? Because the environment provided by Amazon is convenient and extensive, and everything you need is there. When starting research, it is faster and more efficient to use, for example, software that has been developed by your predecessors whenever possible, rather than installing computers and writing each and every program yourself.

Taki: Is this trend common within the scientific community?

Kitsuregawa: It is already standard in genome research in the life sciences. Sequenced DNA data are shared, and researchers, fully aware of what forms the core of their own competitive strength, take whatever they need or think is good from the published data. Google's decision to open up its deep learning library is similar, but its intention was to attract leading researchers and spread Google's methodology. Having people publish libraries and allow anyone to use them freely is one way of doing things, but I think the time will come when we will agree to place everything on the same platform. That is what NII is aiming toward.

Taki: Moves to provide IT platforms that support research are also being made by major publishers and others in the private sector.

Kitsuregawa: The commercial viability of such services is close in industry, and this is an area of intense competition between researchers. Research funds are plentiful, and commercial services are feasible. Personally, I want to focus on supporting fields that commercial services show little interest in. Also, research tends to "effervesce" in interdisciplinary fields. As an inter-university research institute, NII frequently deals with diverse players, and so we will take advantage of this attribute to try to support interdisciplinary research.

(Photography by Seiya Kawamoto)

A Word from the Interviewer

Data on sequenced DNA have been shared by the Human Genome Project, a collaborative project between researchers in countries including Japan, the US, and the EU. It also appears that experimental and observational data from equipment such as giant accelerators and large astronomical telescopes have been shared for some time. There is, without doubt, a major trend toward sharing data.

However, it is hard to believe that unregulated data sharing will proceed in all fields. The situation is not straightforward in fields where there is fierce competition between researchers, companies, or nations. Some data must be protected. I have often heard complaints and concerns about information being leaked at the peer review stage when manuscripts are submitted to European or North American journals. A similar situation in the world of data must be avoided. This will require Japan to actively contribute in the creation of rules on the publication of data.

Junichi Taki

Editorial Writer and Senior Staff Writer, Economic Commentary Department, Nikkei Inc. Joined Nikkei Inc. after graduating from the School of Political Science and Economics, Waseda University. Held positions in the Industry Department (now Business News Department) and the Washington Bureau, and was Senior Staff Writer in the Economics Department at Osaka Head Office, as well as Head of the Science and Technology News Department, Tokyo Head Office, before becoming Editorial Writer in March 2009. Areas of responsibility include science and technology, environment, and medicine.

Forefront of Open Data

LOD and DOIs making the Data Web a reality

Hideaki Takeda
(Professor, Principles of Informatics Research Division, National Institute of Informatics/
Professor, School of Multidisciplinary Sciences, The Graduate University for Advanced Studies)

A system of open data makes it possible to instantly locate research data and articles that are publicly available throughout the world, and to freely link to and make use of this information. What kinds of mechanisms enable open data, and what kinds of work are currently being undertaken by relevant organizations to ensure further convenience? I asked Professor Hideaki Takeda, an expert on this research topic who is also familiar with the situation outside of Japan.

LOD making the Data Web a reality

Data that are available for anyone to use, reuse, and redistribute, that is, a system of open data, are essential in promoting open science. Recently, the Linked Open Data (LOD) have attracted attention and are starting to be put into practical use. Professor Takeda explains, "The LOD concept structures data, including metadata such as publisher and publication date, for computer processing and allows different datasets to be interconnected. It was created with the aim of realizing the so-called Data Web."

LOD use has spread, mainly in Europe and the US, based on the idea that local governments, companies, organizations, and other bodies that disseminate information will release data using a standard format, which will allow the data to be interlinked and enable the Web itself to function as a giant database. Starting with the sharing of experimental data by bioscience research institutions and companies, the creation of databases for bibliographies and sources by libraries, and the provision of regional statistics by local governments, LOD are now being created and used in various fields.

"To give a straightforward example, all of the bibliographies and sources relating to Natsume Soseki that have been published and collected by libraries worldwide have been linked and can be instantly searched and used." (Professor Takeda)

Development of necessary environments for using data

In regard to LOD, the Resource Description Framework (RDF), a method of representing machine-processable Web resource information, and SPARQL, a query language, have been standardized. Information entered into global databases based on RDF is obtained and utilized using applications written in SPARQL.

Currently, in addition to administrative information and public data being made available on portal sites established by the national and local governments of various countries, diverse datasets throughout the world are being cataloged using websites such as "the Datahub", and so it is becoming possible to obtain data. Also, websites such as "Linked Open Vocabularies (LOV)" provide schemas that define data elements based on RDF, and these can be used to build shared databases. Libraries and tools

Hideaki Takeda

for processing LOD are also widely available, and it is becoming relatively easy to build systems from collection to use of LOD.

In Europe and North America, which are leaders in terms of using LOD, a community project called DBpedia that extracts information from Wikipedia and makes it available as LOD is gaining popularity. However, this information is in English, and this creates a barrier to registration and use from Japan. Therefore, NII released "DBpedia Japanese" in May 2012 (Figure).

DBpedia Japanese is implemented as part of the LODAC project run by Professor Takeda.

"The aim of DBpedia Japanese is to provide a DBpedia based on the Japanese language version of Wikipedia. LODAC project has published various datasets as LOD to show that LOD is beneficial for data publishing, for example we collect information on collections in Japanese museums and art galleries and build it as LOD. It is Japan's largest database of museum collections. We also built a database of information on biological species for biodiversity information," says Professor Takeda. Furthermore, mechanisms for collecting and releasing data on common vocabulary infrastructure are steadily being put in place.

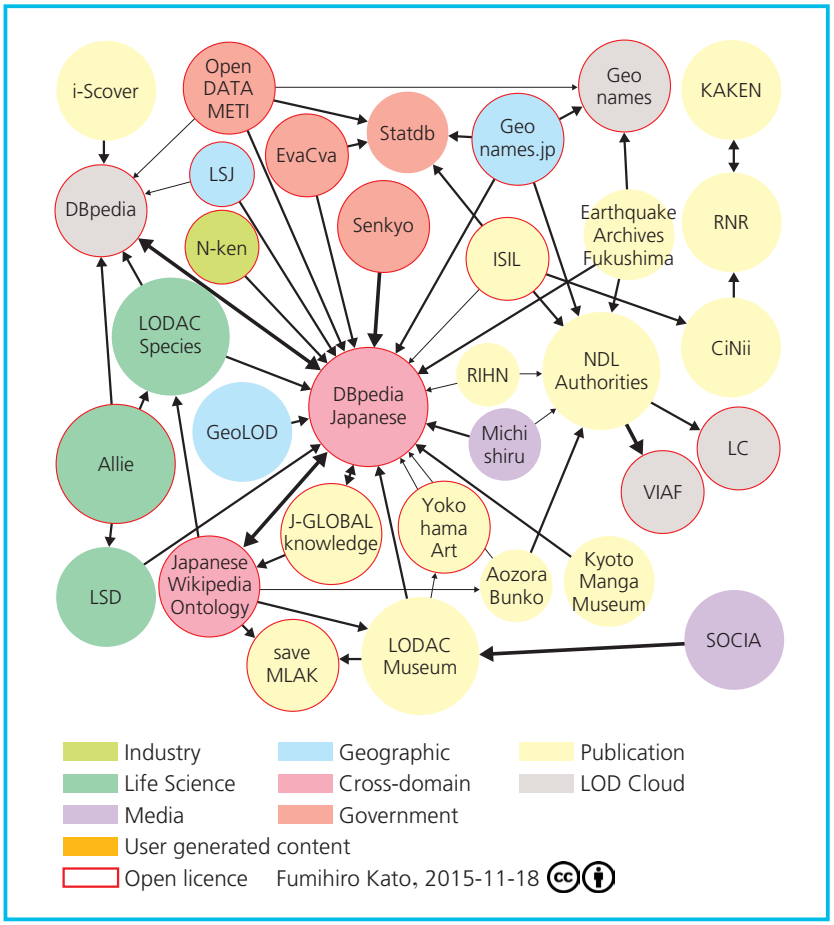


Figure **Japan's Linked Data Cloud**

DOI resulted from digitization of articles

Another initiative for using open data is the digital object identifier (DOI) system. This system makes digital objects on the Internet persistently accessible by adding identifiers to academic papers, as well as registering metadata to make a paper's URL, publication data, publisher, and other information identifiable.

"DOIs were devised jointly by publishers in the 1990s when digitization of academic papers began. If the location of a digitized article is written using a URL, the article may become inaccessible when the URL changes, for example, when a website is updated. Therefore, adding a unique ID to the article itself, unrelated to the URL, makes the article accessible even if the URL changes." (Professor Takeda)

Adding a DOI to an article not only makes it possible to always know its location but also makes it easier to identify citations. Currently, the world's largest DOI registration agency, CrossRef (USA), has registered and assigned DOIs to more than 70.4 million articles worldwide and linked these to citing/cited documents. DOIs have become essential as a common foundation in research.

Meanwhile, partly due to the language barrier, there are only about 1.5 million articles from Japan registered. Therefore, with the aim of popularizing DOI and improving the accessibility and convenience of academic content in Japanese, the Japan Science and Technology Agency (JST), the National Institute of Materials Science (NIMS), the National Diet Library (NDL), and NII established the Japan Link Center (JaLC). JaLC is seeking the participation of every organization handling academic content in Japan, and making efforts to popularize DOI and improve the convenience of information services both domestically and overseas.

"Recent research has been attempting to create promising infrastructure for open science by attaching DOIs not only to articles but also to research data," says Professor Takeda. JaLC is conducting an experimental DOI data registration project together with Japanese research institutions. This project is furthering efforts toward the full-scale use of data DOIs by identifying issues in future system development and operation.

"Whereas LOD are data that anyone can publish, the origin of DOI data is clear and the reliability of the data is guaranteed to some extent. If both types of data are made compatible and freely linked, they will further expand research activities. I want to achieve this by cooperating with people in various fields to solve each problem," says Professor Takeda.

"From now on, not only will research and development relating to open data accelerate, but social structures will change. Before long, both the data and the people who created the data will be directly connected and collaborations will be formed, giving rise to new innovations. This has the potential to transform the existing framework of companies and organizations."

(Interview/Report by Hideki Ito.
Photography by Yusuke Sato)

From Institutional Repositories to Open Science

Kazutsuna Yamaji
(Associate Professor, Digital Content and Media Sciences Research Division/Academic Repository Office, National Institute of Informatics)

Asanobu Kitamoto
(Associate Professor, Digital Content and Media Sciences Research Division, National Institute of Informatics/
Associate Professor, School of Multidisciplinary Sciences, The Graduate University for Advanced Studies

Associate Professor Kazutsuna Yamaji is involved in the Institutional Repositories Program, which is an electronic archive system for collecting, storing, and disseminating the intellectual output of education and research institutions. He is also involved in the operation of the Academic Access Management Federation (GakuNin), which comprises universities and other organizations who use electronic resources. Associate Professor Asanobu Kitamoto is engaged in research projects relating to large databases of global environmental data, the database of old books of the National Institute of Japanese Literature, and the digital archive of Toyo Bunko rare books. The two talked about the current state of development of open science, as well as the challenges it faces and its future prospects.

The key to evaluation standards is trust

—Open science is a very broad concept.

Kitamoto: Open science is viewed differently by different people, and those involved might be regarded as strange bedfellows. Currently, citizen science, open access, open data, collaboration, and crowdfunding are all called open science. To me, open science is meta-research, or research on research. Different people have different approaches to opening up science and different reasons for doing so.

Yamaji: Yes, "open science still hasn't been defined. Everyone thinks that they are onto something, but no one has had a resounding success. Personally, coming from the perspective of open research data, I am developing an open-source system for the NII Institutional Repositories Program and offering this cloud service to various organizations. Japan has the largest number of institutional repositories in the world with more than 500, and of these, more than 200 use NII's repository module, WEKO*. Until now, academic institutions have used it mainly for articles, but articles are only the final output, or the trigger for subsequent research. However, open science makes it easier to incorporate processes into your research. A key point is to work out how to provide infrastructure that supports not only opening up data but also closed areas such as who used the data and what for.

Only NII is capable of tackling this difficult infrastructure building. Furthermore, infrastructure will receive acclaim for being made by a neutral public institution and its user friendliness. In practice, many needs and problems have been identified, but that is partly because there has been so much contact with users. This kind of communication with users generates further trust and helps to create an essential infrastructure and brand. The advantage of a research institute operating the infrastructure is that the research institute can constantly develop new things at the same time.

Kitamoto: However, the question of how to evaluate contributions to infrastructure is a challenge. In the case of the weather database "Digital Typhoon", which I operate, continued updates over the long term have engendered trust from users. However, if the purpose of a database is just writing articles, continued updates will not be subject to evaluation. The question of how to evaluate activities that contribute to the research community by operating data infrastructure openly is an important issue in open science.

Another issue is how to combine open and closed science to generate value. Collaborations arise from sharing research findings, and research costs decrease as a result of reusing data. All of this requires connecting online.

Yamaji: How can research be accelerated simply and conveniently in an IT environment? That is one of the roles of NII. In particular, cloud environments are not widely used in the humanities and social sciences.

Kazutsuna Yamaji

I would like to promote the convenience of these environments.

Kitamoto: One reason might be the fact that a great deal of research in the humanities is individual research. There has never been much collaborative research. But, hereafter, individual research may become increasingly limited, even in the humanities—to say nothing of the fact that collaboration is crucial in the area of "digital humanities", which looks at the role of the humanities in the digital age.

Yamaji: When data are opened up, this attracts the attention of other people, and if there are frameworks in place, new cycles of research begin as people reuse the data. Open data can ignite new research.

Where should the budget come from?

—How about downsides?

Yamaji: There is the problem of whether the people who put out the data are evaluated correctly.

Kitamoto: The assignment of identifiers to research data is being promoted through digital object identifiers (DOIs) and the agency that registers them, Japan Link Center (JaLC). Assigning identifiers to data to specify the creator will make it possible to evaluate their contribution. This relates to the topic of specifying the contribution of authors in the case of articles. Until now, there was only the category of "author" for articles, and so in big science where you have many people involved, there could be as many as a thousand authors. Recently, contributions to research tend to be broken down in more detail, rather than simply "author". But if it becomes possible to measure the contribution of researchers, there is a danger that it will take on a life of its own and be used in unintended ways.

Yamaji: That kind of distortion might instead become a driving force. We are now in an age of guaranteeing transparency by going public.

Kitamoto: This may be where our opinions diverge. Making it possible to store data and check for fraud is an urgent issue, and it is easily understood by the party paying the budget. But if the aim of open science is taken to be preventing fraud, it won't generate much value, will it?

Yamaji: I hold the opposite view. Infrastructure is key, and it is costly. If guaranteeing transparency resonates with investors, then it is okay to use that as a selling point.

Kitamoto: This goes to show that there are various points of conflict in open science, which is why discussions that consider the big picture are necessary.

Cycle of becoming established and gaining trust

—What are the future prospects of open science?

Kitamoto: Various problems are arising in the current way of conducting research. The Internet is the driver that will change this for the better. Behind the emergence of the concept of open science is the view that openness using the Internet has been insufficiently explored.

Yamaji: The question of what researchers can put out from the bottom up is key. We consider our mission to be to create an environment in which it is advantageous for researchers to release research data and lab notes. The raw materials are ready. There are also structures connecting data and cloud infrastructure, as well as authentication mechanisms and repositories. If connected, it might be the start of something new.

Kitamoto: When you create data and make that data publicly available, you get contact from unexpected places, and there are advantages if the data are considered a long-term investment. However, unless the data are created by people or institutions who are in positions that allow them to do so over the long term, it may be difficult to recoup investments.

Yamaji: But if you make your data publicly available, someone will find them.

Kitamoto: It takes time though.

Yamaji: That is why NII is so important as an organization that can build and operate infrastructure over the long term.

Kitamoto: The infrastructure won't win people's trust until it has been maintained for five or ten years. That's the way it is with infrastructure.

Yamaji: Maybe you can't see the appeal of it unless you are someone who is actually providing the service. It is hard work, but you also get the pleasure of seeing the service spread throughout the country.

Kitamoto: Well, I certainly hope that you experience that pleasure along with NII.

(Written by Kazumichi Moriyama. Photography by Mariko Tosa)

Asanobu Kitamoto

*WEKO
A repository system that operates on NetCommons2, developed by NII with the aim of storing academic work and making it accessible. "WEKO" means "repository" in Swahili.

Making Use of Real-World Data in Informatics Research

The role of the Center for Dataset Sharing and Collaborative Research (DSC)

Keizo Oyama
(Professor, Digital Content and Media Sciences Research Division & Director, Center for Dataset Sharing and Collaborative Research, National Institute of Informatics/ Professor, School of Multidisciplinary Sciences, The Graduate University for Advanced Studies)

The application of artificial intelligence technologies, such as deep learning, and big data processing technologies in industry has accelerated in recent years. Meanwhile, society also requires early implementation and industrial application of academic research, and the use of data originating in the real world is becoming more important. Therefore, NII's Center for Dataset Sharing and Collaborative Research (DSC) was established to undertake the mission of making use of large-scale data accumulated in the real world in informatics research. Center Director Keizo Oyama talked about the significance of the Center, which connects researchers and industry, and receives and provides data as a research resource while tackling issues such as protection of intellectual property and privacy.

Keizo Oyama

Mutually beneficial for researchers and corporate providers

Recently, large-scale data processing technologies are regarded as essential for creating new businesses and enhancing services, and there is strong demand for early implementation of relevant research findings. Expectations are particularly high for research fields such as statistical machine learning and big data analysis. The data that researchers have traditionally created by hand for research purposes are inadequate for responding to this societal demand, and the acquisition of large-scale real data from the real world is vital.

Meanwhile, the adoption of the latest and most suitable technologies is becoming a source of competitive strength in industry as, for example, Web-based businesses launch full-scale research organizations and venture companies possessing advanced technologies secure footholds in the market. However, conducting research within one's own organization is insufficient, and there has been an increase in the number of companies wishing to carry out joint research with universities and other public research institutions, even if that means providing company data. Also, by providing data to graduate students doing specialized research, these companies aim to raise awareness of their company and secure talented professionals.

As you can see, researchers and companies have some coinciding interests, but in practice, it is difficult for researchers to negotiate with companies individually. Also, corporate data involve many restrictions, such as confidentiality, copyright, and privacy protection, and complicated terms of use have to be adjusted on an individual basis before the data can be used. Therefore, DSC acts as an intermediary between the two parties, receiving data from companies and providing the data to researchers according to certain rules.

A steady stream of research findings

The impetus for establishing DSC in the first place was the "Yahoo! Chiebukuro" dataset, which Yahoo! Japan Corporation provided for the "NII Testbeds and Community for Information Access Research (NTCIR)", an evaluation workshop started by NII at the end of 1997. This dataset was provided not only to the workshop participants but also to numerous universities and companies, including for research purposes other than those of the workshop, and it was used in a variety of research studies. This attracted attention and a steadily increasing number of companies approached us wanting to provide data to researchers via NII. Therefore, in 2010, NII established the Informatics Research Data Repository (IDR) as a contact point for receiving and providing data. Then in April 2015, it established DSC to further promote data sharing and data use. DSC aims to promote open science centered on data as a research resource by integrating the

Table 1 **Datasets provided by DSC (mostly in Japanese)**

Provider organization	Provided dataset
Yahoo! Japan Corporation	Yahoo! Chiebukuro data (2nd Edition): Q&A data
Rakuten, Inc.	Rakuten Ichiba: product and review data from e-commerce service Rakuten Travel: facility and review data from travel service Rakuten GORA: facility and review data from golf service Rakuten Recipe: recipe data and images from cooking recipe service Rakuten Auction: review and transaction data from auction service Annotated data on some parts of Rakuten data Rakuten Viki: Video information, user behavior and evaluation information
Dwango Co., Ltd. and Brazil Co., Ltd.	Niconico Douga comments: metadata and comment data from video sharing service Niconico Pedia data: article and comment data from collaborative dictionary service
Recruit Technologies Co., Ltd.	Hot Pepper Beauty data: facility and review data from hair salon reservation service
Cookpad Inc.	Cookpad Recipe data: recipe data from cooking recipe service Cookpad Menu data: menu data from cooking recipe service
NEXT Co., Ltd.	HOME'S data: Leased property data and image data from rental housing information service
National Institute of Japanese Literature (NIJL)	NIJL Dataset: Japanese Classic books data (bibliographies, images, tags, and some full texts)
NII	NTCIR test collections Speech corpora Conversation corpora (in preparation; includes speech and image data)

(Note)As of January 12, 2016 (Source) http://www.nii.ac.jp/dsc/idr/datalist.html

operation of NTCIR with the activities of the IDR contact point and NII's Speech Resources Consortium (SRC).

In addition to several tens of datasets such as NTCIR's test collections and SRC's speech corpora, DSC has so far received fourteen datasets from six private-sector corporations and one dataset from the National Institute of Japanese Literature, which it is providing free of charge to researchers in informatics and related fields (Figure 1, Table 1). The data include images, audio, and video, as well as text, and with some exceptions, are available for download via the Internet.

Looking at the use of these datasets, the number of users (e.g., laboratories in universities) has grown steadily from 2007, before IDR was established, to reach 468 at the end of November 2015 (Figure 2), and the number of papers reporting the results of research using these datasets had grown to 350 by the end of 2014, which are both part of accelerating upward trends.

The research results are truly diverse. They include, for example, automatically interpreting cooking recipe data and creating flow diagrams in which multiple tasks are carried out at the same time, seeking optimal solutions to problems from Q&A data, and estimating the chorus or "hook" of a piece of music from video comment data.

Resolving issues of data protection using cloud computing

I hope that more and more highly practical research will come out, but this will require still more diverse data. The biggest issue here is how to protect potentially private and confidential information which may be included in the data. For example, even if explicit personal data are not included in the provided datasets themselves, private information is sometimes revealed when data are matched with other data. In fact, there was an incident in which the release of search query data by a well-known US search engine company led to a user being identified and her privacy violated. This incident still deters companies from providing data.

Currently, IDR negotiates with data providers to have defined conditions of use in advance followed by exchanging memorandums with data users accordingly, as well as taking other measures, but we would eventually like to introduce a system that allows data to be used safely in a cloud computing environment. Various methods of providing data are being considered including usage constraints such as preventing downloading, providing API returning statistically processed results only, and making data accessible only from programs that users create and register.

In order to facilitate the use of datasets while protecting data, it is essential to utilize technologies based on the use of the cloud to strike a realistic balance between the safety sought by companies and the methods of use desired by researchers. To that end, we are currently conducting joint research with companies, and I hope to put some of the concrete results in operation next year.

(Written by Masahiro Doi.
Photography by Yusuke Sato)

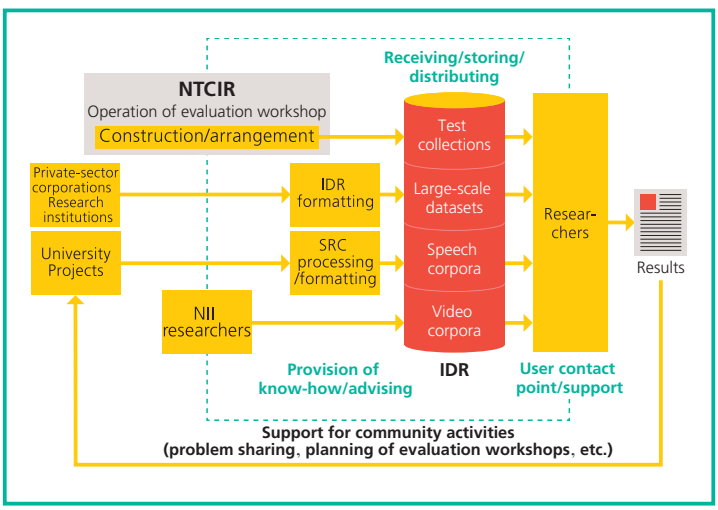


Figure 1 **Activities related to dataset provision by DSC**

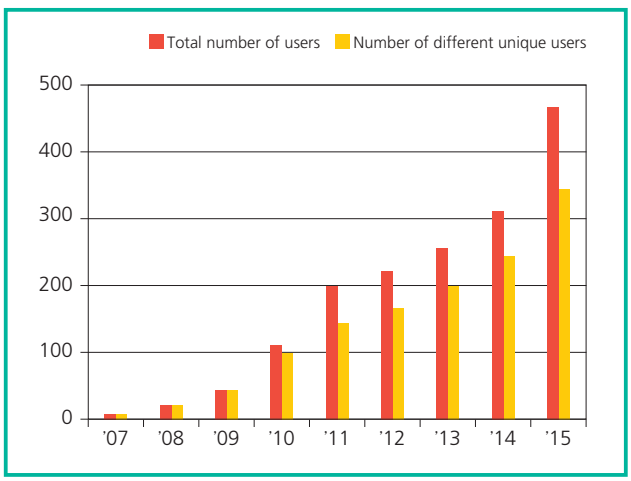


Figure 2 **Trend in cumulative number of dataset users (datasets provided by private-sector corporations: excluding Niconico datasets)**

News 1 Discussions on Open Data
—Symposium held by ROIS

The Research Organization of Information and Systems (ROIS) held a symposium titled "Opening Up Research Data in the Context of Open Science" on February 8.

Setsuo Arikawa (former President of Kyushu University), Chairman of the Cabinet Office's Expert Panel on Open Science discussed the necessity of and problems involved in the international trend that is open science in a keynote speech titled "New Developments in Science due to Opening Up." He also proposed the use of the shared repository service JAIRO Cloud, appealing to participants by saying, "You can release your own papers and data starting today. Let's begin with universities."

This was followed by lectures on "Polar Science and Open Data" by Dr. Yasuhiro Murayama, Director of the NICT Integrated Science Data System Research Laboratory, and "Life Sciences and Open Data" by Professor Toshihisa Takagi, University of Tokyo. Yoko Shintani, Open Research Marketing Manager of Nature Publishing Group, introduced Scientific Data, which has developed research data sharing into a service, and emphasized the importance of strengthening incentives for releasing data.

In the second half of the symposium (photograph), Professor Hiroshi Maruyama and Professor Satoshi Yamashita of the Institute of Statistical Mathematics, Professor Satoshi Imura of the National Institute of Polar Research, Associate Professor Tsuyoshi Koide of the National Institute of Genetics, Project Associate Professor Mari Minowa of the Database Center for Life Science, and Associate Professor Asanobu Kitamoto of NII's Digital Content and Media Sciences Research Division discussed "Promoting Open Data in Research Settings." They indicated future directions, saying, "Rather than distinguishing between articles and data, we should evaluate contribution to science," "Releasing failed data that do not become evidence for an article will create the potential for innovative research," and "Developing environments that facilitate data release and raising awareness are important."

News 2 Last Industry–Government–Academia Collaboration Lecture of the Year Reporting the frontline of research on "material perception"

The 5th Industry–Government–Academia Collaboration Lecture, the last one of this fiscal year, was held on January 22. The lecture, titled "Development of Research on Material Perception," was given by Professor Imari Sato (photograph) of the Digital Content and Media Sciences Research Division, who specializes in computer vision.

She first explained "how we see things" by analyzing spectral characteristics, showing that how we see something changes completely according to differences in light source and whether there are reflections. She then described the state of progress of research into methods for consistently estimating objects from images by presenting sampling methods that allow objects to be reproduced using few samples. Also, taking CG as an example, she described how results from material perception research are being used in creative fields.

Professor Sato appealed for active exchange of information, saying, "If people in industry share their research and development to the full extent possible, the topics of research available to us researchers will expand, and it will be mutually beneficial."

News 3 Exploring the True Nature of Bubbles Using Big Data
—5th NII Shonan Meeting Special Lecture

The aim of the NII Shonan Meetings is to advance informatics by assembling the world's top researchers in the field to discuss currently unresolved problems and seek solutions. The 5th NII Shonan Meeting Special Lecture was held on December 13 as an outreach activity. The theme was "Understanding the Current State of Economics: The True Nature of Bubbles Explored Using Big Data." The lecture was given by Associate Professor Takayuki Mizuno of the Information and Society Research Division, an expert in econophysics.

Professor Mizuno suggested that "the keyword of bubbles is 'disparity'," and gave real estate data as an example. He explained that whereas the size and value of real estate properties are normally roughly proportional, during bubbles the value of specific properties even under the same conditions soars for speculative reasons, resulting in "variability". In the case of stock prices, too, he said that it is possible to judge whether stock price rises are the result of a bubble or sustained economic growth by observing disparity between brands using big data.

He also introduced recent studies such as predicting the global financial crisis through economic networks and economic forecasts using the news and Twitter. Attendees heard about attempts to solve familiar problems using the latest research.

Flash

▶ **4th SPARC Japan Seminar 2015**

Held on March 9, titled "The Function of University Libraries in the Context of Promoting Research." Koichi Ojiro (University of Tokyo Library System), Takashi Hikihara (Director-General of Kyoto University Library Network), Hiroshi Manago (Cabinet Office), and Setsuo Arikawa (former President of Kyushu University) spoke about open access, open science, and the role of university libraries. Chaired by Nami Hoshiko (Kyushu University Library).

In the second half, there was a panel discussion on the subject of how university libraries can contribute to improving Japan's research capabilities, moderated by Midori Ichiko (Hiyoshi Media Center, Keio University).

Topics

Focusing on Financial Smart Data and Cognitive Technology with the Establishment of Two Research Centers as Bases for Spreading Innovations

NII promotes collaboration between academia and industry, and on February 1 it established two new research facilities. The Financial Smart Data Research Center, established jointly with Sumitomo Mitsui Asset Management Co., Ltd. (SMAM), was announced at a press conference on February 9. Six days later, establishment of the Cognitive Innovation Center was announced. Research at this center will be supported by IBM Japan, Ltd.

Research facilities are departments devoted to research in specific fields, and the establishment of these two centers brings the number of NII research facilities to eleven. The purpose of both centers is to give research findings back to society, and their aim is to nurture the roots of innovation in society rather than strengthening specific technological capabilities.

The head of the Financial Smart Data Research Center is NII Director-General Masaru Kitsuregawa. NII used the Joint Research Department System just introduced in February by the Research Organization of Information and Systems (ROIS), which uses private-sector money to establish and operate research departments that have a high level of public benefit. This is the first time that NII has used private-sector money to establish a research facility.

Financial smart data are created by taking big data, which are merely a huge collection of complicated data, and processing and analyzing them to transform the data into useful knowledge that will lead to new value creation. By using financial smart data to try to understand the laws of economic/social phenomena, this center aims to implement long-term future predictions and, by extension, to fulfill the social mission of stimulating domestic financial markets and creating stable national wealth.

Director-General Masaru Kitsuregawa and Kunio Yokoyama (left), President & CEO of SMAM, announcing the joint establishment of the Financial Smart Data Research Center.

(From right) Center Head Mitsuru Ishizuka, Director-General Masaru Kitsuregawa, and Cameron Art of IBM Japan at the press conference for establishment of the Cognitive Innovation Center.

Meanwhile, former president of the Japanese Society for Artificial Intelligence Mitsuru Ishizuka (Professor, Waseda University; Professor Emeritus, University of Tokyo) was invited to head the Cognitive Innovation Center. The central theme of "cognitive technology" includes machine learning, natural language processing and understanding, and intelligent information processing such as constructing and using big data and knowledge bases. The center will focus on supporting human cognition and judgment in natural interaction by extensively applying not only the latest artificial intelligence technologies, such as deep learning, but also advanced information technologies in order to learn from big data.

Many of Japan's leading companies from a wide range of industries will participate in the activities of this center. The aim is to bring about two transformations: a radical change in thinking towards promoting the application of cognitive technology in society, and the discovery of new links between state-of-the-art technologies and industry.

SNS

"Hey, this is great!"

Hottest articles on Facebook and Twitter (Dec 2015–Feb 2016)

National Institute of Informatics, NII (official) Facebook
www.facebook.com/jouhouken/

[Hi! from Bit-kun] LOVE Bit
Today is Valentine's Day!
Bit sends you all lots of love!
(February 14, 2016)

National Institute of Informatics, NII (official) Twitter
@jouhouken

[NII News] Assistant Professor Takuya Akiba received the 2015 Funai Research Incentive Award.
(February 28, 2016)

Bit on Twitter! Twitter
@NII_Bit

The conversation between Professor Noriko Arai and Daisuke Tsuda broadcast on New Year's Day on Jam the World New Year Special #jwave has been published in Daisuke Tsuda's e-zine @tsuda! Bit
(January 27, 2016)

Essay

A World in Which Scientists are Cool

I was asked recently about my favorite SF, which led me to recall the SF I loved to read during my junior high school and high school years and also to wonder, "What is the definition of SF?" There seem to be various opinions on this, but since the "S" stands for "science", perhaps SF is fiction that depicts a future world in which science has advanced? Or, novels that use science as a kind of prop or device? The same kind of ambiguity exists in the difficulty of defining "open science". Expectations and dreams regarding science intermingle, and the boundaries are not easily fixed.

Open science contains wide-ranging concepts. Part of open science might be summarized as "innovative infrastructure for accelerating science". This includes solutions to structural problems in various fields. The obstacles that exist in each field of science are many and varied, and so the aims of open science range widely, from reducing the costs involved in publishing academic journals to citing and assigning IDs to data, building permanent archives, and promoting citizen science. If open science becomes a huge movement that anyone can participate in, it will lead to major innovations.

However, the phrase "open science" has only just begun to be used. If you carry out an image search using this phrase and compare the results with image search results for "big data" or "crowdsourcing", the difference is clear. The majority of the retrieved images are slides full of text, and no common visual images can be found. "Open science" does not seem to be a key phrase that appears frequently in the news right now, maybe because scientific activities, in general, are not so strongly related to people's everyday life. This makes people feel the term is a bit abstract and distant from their familiar world.

Returning to the subject of SF, my own definition is fiction that features cool scientists. The heroes and heroines are clear-headed and they tackle difficult problems using their excellent skills in analyzing information and accurately assessing situations, which is very cool. Open science demonstrates the world in which scientists are active, and so I would like image searches to bring up pictures of cool, stylish scientists!

Akiko Aizawa
Professor, Digital Content and Media Sciences Research Division, National Institute of Informatics

Future Schedule

May 25–27 | Academic Information Infrastructure Open Forum 2016, National Institute of Informatics, Hitotsubashi Hall.

May 27–28 | NII Open House 2016 (public exhibition/presentation of research results), Hitotsubashi Hall. For details and to apply for events that require advance registration, please visit the URL below.
http://www.nii.ac.jp/openhouse/

June 22 | 2016 Public Lectures "The Front Line of Informatics"; 1st Lecture (Takuya Akiba, Assistant Professor, Principles of Informatics Research Division): A program of lectures in which the Institute's researchers explain topics in advanced informatics to the general public, held six times a year. Details for 2016, including schedule and topics, will be announced via the URL below once decided.
http://www.nii.ac.jp/event/shimin/

Notes on cover illustration Depicting robots doing experiments in separate places, this cover represents the world of open science that is realized by sharing experimental data, images, and other information. It suggests a new approach to science offered by collective intelligence.

Weaving Information into Knowledge
NII

National Institute of Informatics News [NII Today]
No. 57 Mar. 2016
[This English language edition NII Today corresponds to No. 71 of the Japanese edition.]
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
Publisher | Masaru Kitsuregawa **Editorial Supervisor** | Ichiro Satoh
Cover illustration | Toshiya Shirotani **Copy Editor** | Madoka Tainaka
Production | MATZDA OFFICE CO., LTD., Athena Brains Inc.
Contact | Publicity Team, Planning Division, General Affairs Department
TEL | +81-3-4212-2164 **FAX** | +81-3-4212-2150 **E-mail** | kouhou@nii.ac.jp

Bit (NII Character)

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