FUJIWARA, Ikki

Associate Professor, Information Systems Architecture Science Research Division

National Institute of Informatics

[MAIN BODY]

■ Building data processing environments is a barrier for researchers

All modern research activities can be described as actions that extract knowledge from data. The paradigm of "open science," which aims to accelerate data-driven research by making available academic papers and supporting evidence as a fundamental principle, is permeating all academic fields. With the Research Center for Open Science and Data Platform (RCOS) playing a leading role, NII created and continues to develop a research data platform, the NII Research Data Cloud (RDC), capable of managing, publishing, and searching papers and research data. NII provides this platform to universities and research institutes throughout Japan. Since taking up my post in 2018, I have been working on the development of a data analysis platform that will serve as an execution platform for various extensions to the NII RDC.

Data are only meaningful after being processed by a computer. It is impossible for researchers to find meaning in vast databases by looking at the data piece by piece. This means that they need a thorough knowledge of both data and computing in order to use public data and practice open science. Services such as the NII RDC are becoming available, making it easy for researchers to find and use data, even if they are not an expert in the target field. In computing, however, researchers must develop or acquire software for data processing, procure user rights for hardware suited to the nature of the processing, and build appropriate execution environments so that the target software can run on the hardware. This sequence of tasks is difficult even for experts in computer systems. I considered this to be a large barrier to using public data and practicing open science for many researchers who are not computer experts.

■ Data analysis function that allows all researchers to focus on knowledge discovery

To begin with, I developed a "data analysis function" as an extension to the NII RDC's data management platform, GakuNin RDM. This function allows researchers to deploy a cloud-based data analysis environment at the touch of a button from the project screen and immediately start writing analysis programs for data managed on GakuNin RDM. This frees them from the onerous task of preparing a program execution environment themselves. It also makes it easy for a teacher, for example, to replicate the created execution environment for many students. In response to requests from researchers, I have also developed a "bring-your-own computer function" that allows computers owned by universities and research institutes to be used instead of the cloud. I am currently working on a "computational reproduction package function" that will allow findings obtained using the data analysis function to be

published together with the data and analysis programs. Providing these functions along with GakuNin RDM to universities nationwide will remove the barrier of needing a thorough knowledge of both data and computing, encouraging researchers in various fields to use public data and practice open science.

In my view, it is NII's duty to provide all researchers with the opportunity to participate in data-driven science. To fulfill this, it is my goal to make the barrier to using data on a computer as low as possible by connecting the world of data and the world of computing—with GakuNin RDM at the core. I want to provide an environment that allows researchers and students in all disciplines and with varying levels of expertise and proficiency to focus on data analysis that leads to knowledge discovery, without worrying about preparing or managing computers.

■ Envisioning next-generation platform services with GakuNin RDM at the core

In January 2024, I transferred from the RCOS to the Center for Cloud Research and Development (CCRD). Looking ahead, advanced functions such as secure computation, data governance, data provenance, curation, and secure storage environment functions will be implemented on the NII RDC. The data analysis platform I worked on will be used as an execution platform for these advanced functions. I will also develop a service that integrates GakuNin RDM with SINETStream, a data collection platform for the Internet of Things (IoT) being developed and provided by the CCRD, to offer comprehensive support—from the analysis of collected data to the storage, sharing, and use of analysis results.

As data-driven science becomes more common, research data will become increasingly diverse and large in volume. I hope to work on developing technologies consistent with the needs of researchers, such as a secure computation function that researchers can feel confident using for sensitive data containing personal or corporate information, and a mechanism that allows GakuNin RDM to handle big data that would normally be analyzed using a supercomputer.

We at the RCOS and the CCRD are participating in the Research Data Ecosystem Development Project promoted by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) as core members of the Research Data Platform Advancement Team. We hope to contribute to the transformation of research activities through the use of data (Research DX) by providing advanced platform services.

Researchmap   https://researchmap.jp/ikki?lang=en

Written by SATO Naoki, translated by FORTE Science Communications.
Updated January 2024