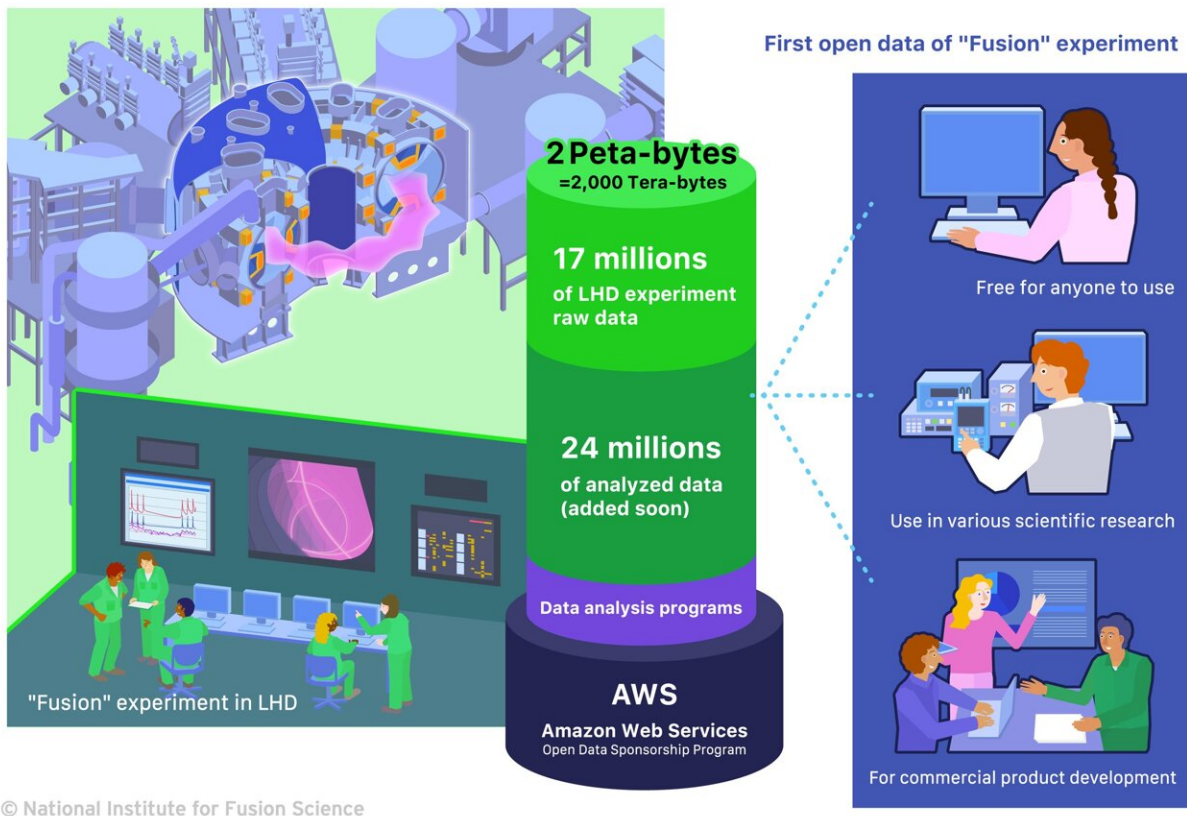


# 25 years of massive fusion energy experiment data open on the 'cloud' and available to everyone

June 14 2024



All diagnostic and analyzed data of LHD, having more than 40 million items and 2 petabytes in total, are open to the public on AWS's cloud storage. Credit: National Institute for Fusion Science

High-temperature fusion plasma experiments conducted in the Large Helical Device (LHD) of the National Institute for Fusion Science (NIFS), have renewed the world record for an acquired data amount, 0.92 terabytes (TB) per experiment, in February 2022, by using a full range of state-of-the-art plasma diagnostic devices.

The International Thermonuclear Experimental Reactor (ITER), which is currently under construction in France through the [international collaboration](#) of seven parties, is expected to generate approximately 1 TB of data per experiment in 10 years, and LHD is currently the only experiment in the world that produces data closely aligned to ITER.

The promotion of "Open Science," in which large-scale research data assets are utilized and shared across society, was adopted as a joint statement at the G7 meeting held in Sendai, Japan in 2023. NIFS started full-fledged efforts toward Open Science by establishing the "Open Access Policy" in February 2022 and the "Research Data Policy" in October 2022.

Since 2023, all the data obtained from LHD experiments are open to the public immediately after acquisition and analysis is completed. All computing program source codes for data analysis are also openly available.

In Open Science, the FAIR Principle is regarded as an important indicator. NIFS considers the fulfillment of the FAIR requirements in diagnostic raw and analyzed data, i.e., valuable digital assets of the LHD project, to be an important proposition of the LHD Academic Research Platform and continues its efforts.

Although LHD experiment data has become one of the world's largest data assets and is widely used by domestic and international fusion plasma researchers, it has been seldom used for other purposes such as in

different research fields or in industry. This may be due to 1) the difficulty of finding the data of interest from a wide variety of experiment data, and 2) the enormous number and the huge size of individual data, which make it difficult to start data analysis easily and quickly.

In order to solve these problems, it is expected that 1) a comprehensive, bird's-eye view of huge amounts of experiment data are enabled, and 2) the data-analysis environment can be easily prepared to start analyses instantly, and data computing resources can be increased or decreased as necessary.

## **Research achievements**

LHD experiment data is a large-scale digital asset. To promote its use by researchers in different fields, industry, and the general public, a computer environment that can be easily used by anyone is necessary. An important possibility exists in "cloud services" technology.

Cloud services provide an environment in which data analyses can be started immediately, enabling researchers, industry, and even citizen users to make use of data very effectively. Now, NIFS has been adopted for the "Amazon Web Services (AWS) Open Data Sponsorship Program", and has completed the data transfer of about 2 petabytes of LHD experiment data onto AWS's cloud storage, [Amazon Simple Storage Service](#) (Amazon S3), to make them freely accessible to anyone on the Internet.

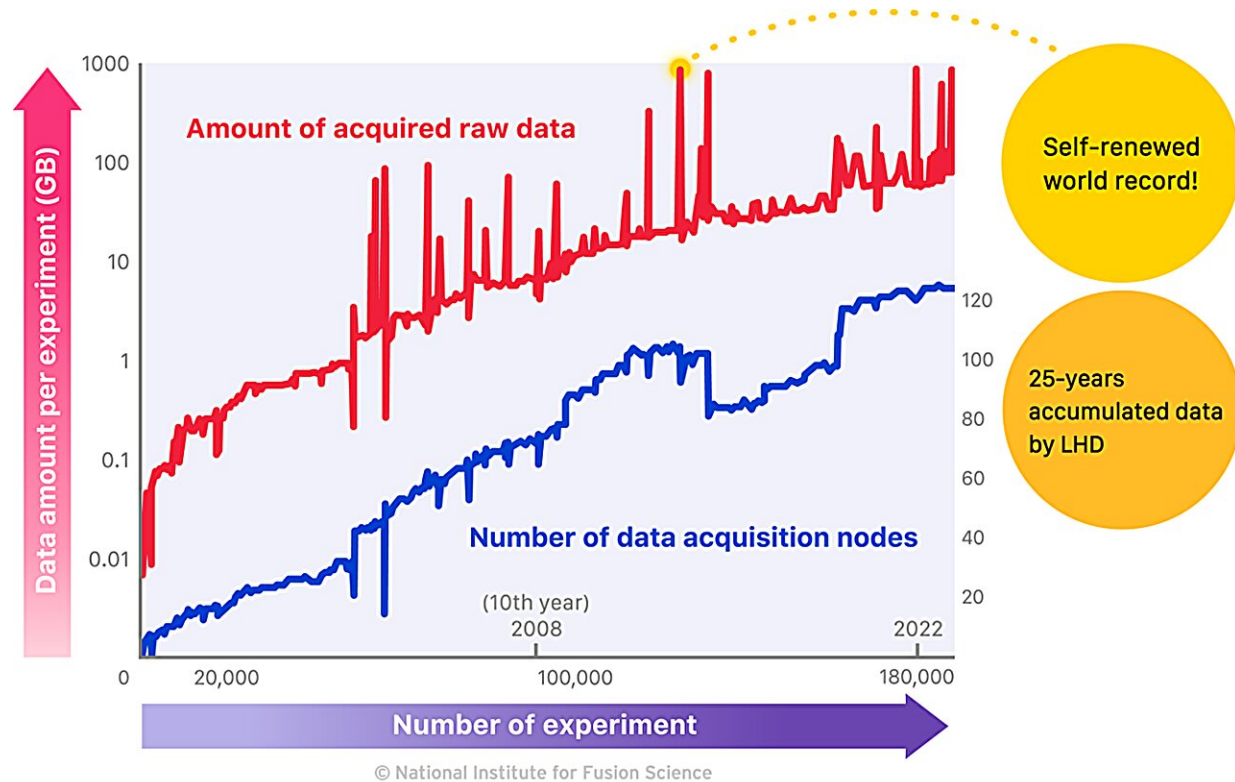
A computing environment capable of running a suite of data analysis programs is also indispensable for the utilization of vast open data. LHD data replicated entirely on AWS's cloud storage can now be accessed directly from AWS cloud computers for high-performance, massive data analyses at any time.

It is also a major advantage for the promotion of Open Science that Amazon S3 enables us to provide a reliable, nonstop data service, independent of the NIFS system and network capabilities.

Unlike other research fields, such as global environmental, meteorological, and astronomical observations, where international research data sharing has already been taking place for more than a few decades, there has been little international data collaboration or sharing in fusion energy research and development, especially in the experimental field.

This is because experimental results often differ from one device to another, making it difficult to simply compare and evaluate them. The LHD open data represents the world's first major step towards interdisciplinarity and universalization of fusion energy research.

The results will be presented at the [14th IAEA](#) Technical Meeting on Control Systems, Data Acquisition, Data Management and Remote Participation in Fusion Research to be held in São Paulo, Brazil, July 15–19, 2024.



Amount of acquired raw data per LHD experiment (red line) and the number of diagnostic devices (blue line). Credit: National Institute for Fusion Science

### Significance of achievements and future developments

The LHD diagnostic raw and analyzed database, which is the world's largest accumulation of fusion energy research data, is a very valuable digital research asset. By making all of it as open data on the AWS cloud, it is expected that the database will not only be used for research purposes within and outside fusion research, but will also attract participation from the general public and new entrants from other countries and industries that wish to start new fusion energy research and development.

The barriers for first entry are expected to be lowered significantly. In addition, it is expected to be a major digital platform for research knowledge exchange, human exchange and development not only in Japan but also elsewhere in the world. For this purpose, NIFS intensively promotes this large data repository under the name of the "Plasma and Fusion Cloud", by using the NII RDC, the research data cloud platform of the National Institute of Informatics.

In the future, to advance Open Science principles, we have just started assigning a global persistent identifier, DOI (Digital Object Identifier), to about 40 million LHD data to facilitate their findability and accessibility. It may take three to four years to complete registration, due to the extremely large number of data entities.

However, when all the data is registered, it is expected to be the largest number of publicly available research data DOIs in the world, exceeding the current world leaders such as Geoscience Australia (approximately 7 million DOIs), CERN (approx. 6.7 million), and the Interdisciplinary Earth Data Alliance (IEDA) in the U.S. (approx. 5 million).

Ushio Usami, the country leader for AWS worldwide public sector in Japan, said, "We are very pleased to be able to contribute to the utilization of fusion energy in collaboration with the National Institute for Fusion Science. I hope that this open data will be utilized not only in the academic research field in Japan, but also by industries around the world to promote technological innovation in various scientific fields."

Dr. Keiichi Nakano, Chief Researcher for Cyber Science Infrastructure of Research Center for Open Science and Data Platform, National Institute of Informatics (NII), and also the Program Manager of the "Developing a Research Data Ecosystem for the Promotion of Data-Driven Science," said, "In this achievement, the research data infrastructure (NII Research Data Cloud: NII RDC) that we have built

was used as a function for utilizing huge amounts of data.

"We are delighted that the NII RDC was able to contribute to the practical implementation of Open Science, which will have a global impact. We hope to continue to deepen our collaboration with NIFS and contribute to the development of global Open Science through this research data."

**More information:** [National Institute for Fusion Science releases approximately 2 petabytes of data from 25 years of Large Helical Device \(LHD\) experiments as open data on AWS.](#)

Provided by National Institutes of Natural Sciences

Citation: 25 years of massive fusion energy experiment data open on the 'cloud' and available to everyone (2024, June 14) retrieved 26 June 2024 from <https://phys.org/news/2024-06-years-massive-fusion-energy-cloud.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.