

"Plasma and Fusion Cloud"

Data Analysis Environment

M.Emoto, H.Nakanishi, M.Ohsua, S.Imazu, M.Yoshida, M.Nonomura, and R.Sakamoto
National Institute for Fusion Science
emoto.masahiko@nifs.ac.jp

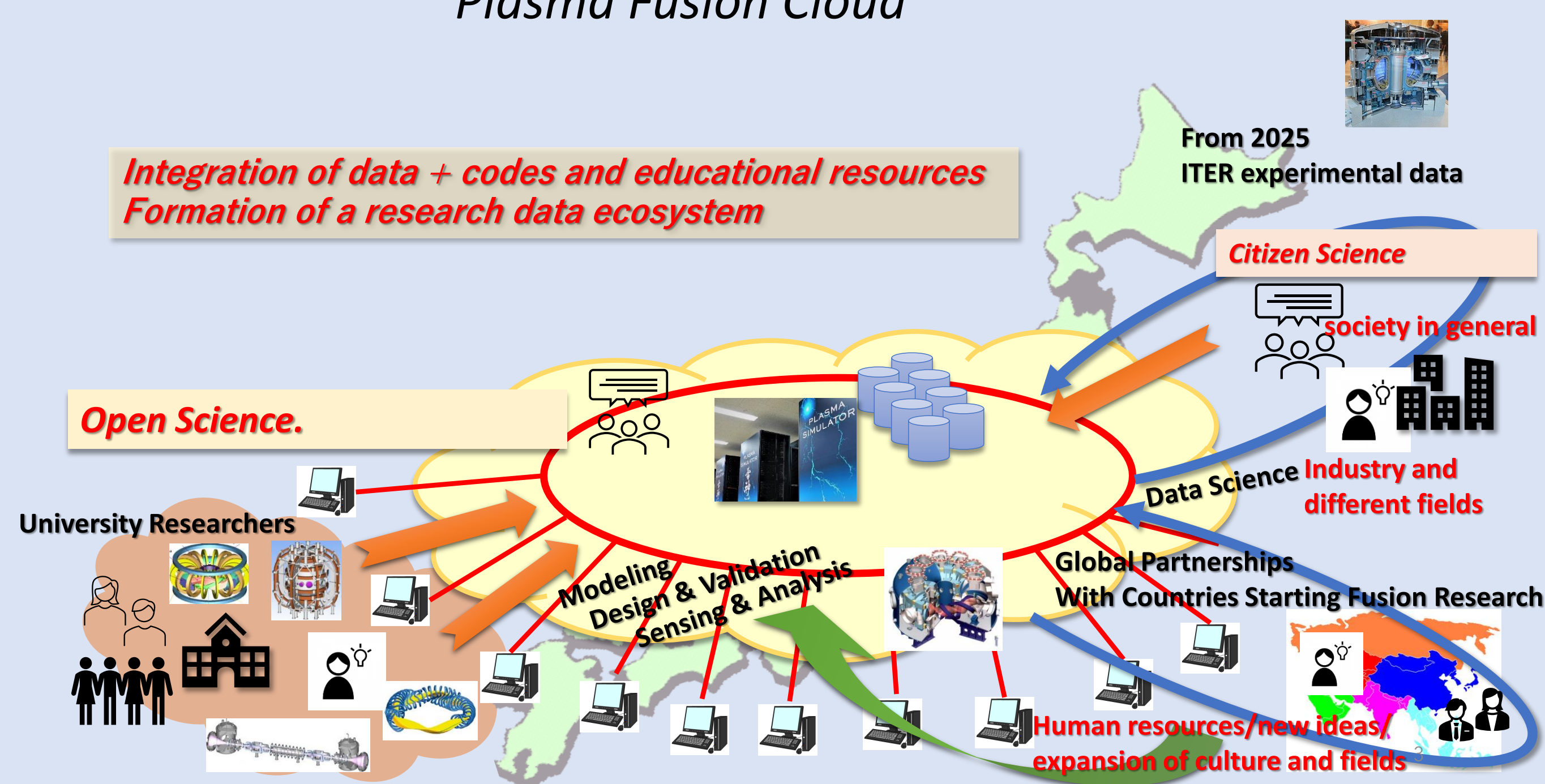
ABSTRACT

- The "Plasma and Fusion Cloud" concept is underway at the National Institute for Fusion Science (NIFS) to create an environment for data analysis that transcends the boundaries of fields.
- The open data server, which is a part of the concept, is now making all physics data obtained from LHD experiments available to the public without delay, and the raw data is also available to the public.
- The data analysis environment is also being developed, including the packaging of the analysis environment and the construction of a data analysis environment using supercomputers and cloud computers such as Amazon Web Services (AWS).

BACKGROUND

At NIFS, the Fusion Cloud concept is underway, and participation in the Fusion Cloud will reduce the burden of data management at each laboratory by managing and providing data obtained from each experimental device in a unified manner. In addition, by providing these data in an open environment on the Internet, it can be expected to be used not only by those involved in fusion research, but also widely in related fields such as plasma physics and condensed matter physics, as well as in the promotion of research in information science as big data. In addition, by providing a data analysis environment in a cloud environment, etc., data analysis can be performed immediately without installing necessary software individually.

Plasma Fusion Cloud



OPEN DATA

NIFS has been fully opening LHD measurement and analysis data to the public (open data) since the 2022 fiscal year. The open data includes raw data, physical data, analysis program codes, instrument overviews, shot summaries, visualization tools and other related utilities, etc. Data are opened immediately after collection and analysis are completed, with no embargo period. Data are available from the LHD Experimental Data Repository without limitation, and total number of data is about 40 million measurements and analysis data, and total data size is about 2 PB.

Raw data is also available from Registry of Open Data on AWS. The program was adopted by the Open Data Sponsorship Program, a social contribution program of Amazon, and approximately 2 PB of LHD experimental data was stored and released free of charge on AWS S3 storage.

EXPERIMENT DATA IS AVAILABLE ON THE INTERNET

Search for Registration data by Shot number

Exp. Code	Shot No.	Exp. No.	Start Date	End Date
4	2225	4322	2002-04-01	2002-02-07
7	4123	4022	2002-04-04	2004-01-21
8	4423	3624	2004-04-17	2004-01-20
9	7022	4022	2002-04-04	2004-01-21
10	8024	7022	2002-04-02	2002-02-19
11	7022	8024	2002-04-02	2002-02-19
12	8024	8024	2002-04-02	2002-02-19
13	8024	8024	2002-04-02	2002-02-19
14	8024	8024	2002-04-02	2002-02-19
15	10424	11247	2011-07-27	2011-10-20
16	10424	10424	2011-07-27	2011-10-20
17	11743	12424	2012-04-02	2012-12-26
18	12424	12424	2012-04-02	2012-12-26
19	12424	12424	2012-04-02	2012-12-26
20	14424	13248	2002-04-01	2002-02-08
21	13247	13247	2002-04-01	2002-02-08
22	14424	13248	2002-04-01	2002-02-08
23	17027	17027	2021-01-04	2021-01-17

Shot number: 171000

Diagnosis	Subject	Comment	Actual Method	Rate
171000-DIAG	1	Reference data of the plasma resist.	reference	
171000-DIAG	2	Reference data of the plasma resist.	reference	
171000-DIAG	3	Reference data of the plasma resist.	reference	
171000-DIAG	4	Reference data of the plasma resist.	reference	
171000-DIAG	5	Reference data of the plasma resist.	reference	
171000-DIAG	6	Intensity of visible light at an upper divertor area viewed from 3-D (14225) telescope	reference	
171000-DIAG	7	Intensity of visible light at an upper divertor area viewed from 3-D (14225) telescope	reference	
171000-DIAG	8	Intensity of visible light at an upper divertor area viewed from 3-D (14225) telescope	reference	
171000-DIAG	9	Intensity of visible light at an upper divertor area viewed from 3-D (14225) telescope	reference	
171000-DIAG	10	Reference data of the plasma resist.	reference	

Registry of Open Data on AWS

The Registry of Open Data on AWS is now available on AWS Data Exchange. All datasets on the Registry of Open Data are now downloadable on AWS Data Exchange at a rate of 1,000 records per second. Data products from category-leading data providers across industries. Explore the catalog to find open, free, and commercial datasets. [Learn more about AWS Data Exchange](#)

NIFS Large Helical Device (LHD) Experiment

Description
The Large Helical Device (LHD), owned and operated by the National Institute for Fusion Science (NIFS), is one of the world's largest plasma confinement devices which employ a heliotron magnetic configuration generated by the superconducting coils. The objectives are to conduct academic research on the confinement of steady-state, high-temperature, high-density plasmas, core plasma physics, and fusion reactor engineering, which are necessary to develop future fusion reactors. All the archived data of the LHD plasma diagnostics are available since the beginning of the LHD experiment, started on 11th of March, 1995.

Update Frequency
Archived data files are updated nightly when new or revised data are generated in LHD experiment.

License
This data is available for anyone to use under the "Rights and Rules".

Documentation
https://www.lhd.nifs.ac.jp/ohd/Repository_en.html

Managed By
NIFS
See all datasets managed by NIFS.

Contact
For any questions regarding data delivery or any general questions regarding the LHD Experiment data repository, please send email to the Data Acquisition and Analysis Group at Data_Analysis@nifs.ac.jp.

How to Cite
NIFS Large Helical Device (LHD) Experiment was accessed on https://www.lhd.nifs.ac.jp/ohd/Repository_en.html from <https://registry.opendata.aws/nifs-lhd/>

Usage Examples
Tutorials
• Data handling in the LHD experiments - Guide for the collaborator (pdf) by Satoshi Okada

Tools & Applications
• How to use diagnostic raw data and/or primary processed data using "Viewer" by NIFS LAOCHI

Publication
• Data Acquisition and Management System of LHD by Hideo Nakashiki, et al.

Open data server

<https://exp.lhd.nifs.ac.jp/opendata/LHD/>

Registry of Open Data on AWS

<https://registry.opendata.aws/nifs-lhd/>

PORTABLE DATA ANALYSIS ENVIRONMENT

In order to verify the effectiveness of the analysis program in a highly scalable environment using cloud infrastructure such as AWS, by connecting the supercomputer system Raijin of the NIFS and the raw data management system via a high-speed network, we created an environment in which analysis programs can be executed on the supercomputer.

In addition, we are working on packaging the analysis environment so that the analysis can be performed immediately not only on such a large-scale system but also on various computer environments used by users and are developing and evaluating the following environments for this purpose.

Container

It is envisioned that data management systems and analysis programs will be executed in a scalable and extensible environment using large clusters and other methods using container technology such as Docker.

Virtual Environment

This method uses Python's Virtual Environment and does not require special permissions for installation, etc., compared to Docker, etc. It is intended for users who want to acquire and visualize data immediately and for relatively light purposes.

CONCLUSION

- The "Plasma and Fusion Cloud" concept is underway at NIFS.
- LHD Experiment is now available from the Open Data Server as well as Registry of Open Data on AWS.
- In order to provide analysis program for various computer, the software is migrating to supercomputer and packaging software to make a portable analysis environment.

ACKNOWLEDGEMENTS

The authors wish to thank all the people who supported this project.