

Progress of “Plasma and Fusion Cloud” Research Data Ecosystem Orienting “Open Science”



***Nakanishi H., Emoto M., Takayama A., Yamamoto T., Yamanaka K.[†],
Urushidani S.[†], Ohsuna M., Imazu S., Yoshida M., Nonomura M.,
Satake S., Todo Y., Osakabe M., Sakamoto R.***



National Institute for Fusion Science (NIFS)

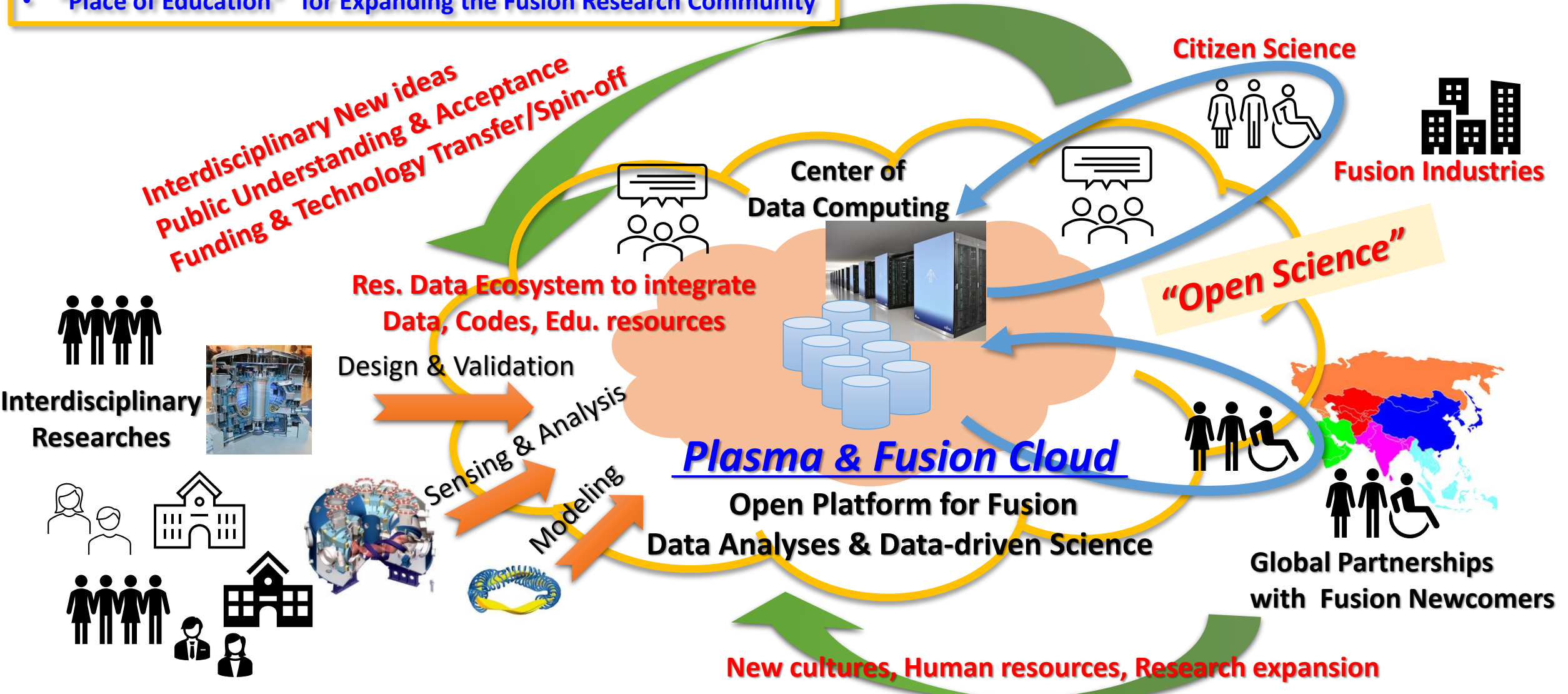
[†]National Institute of Informatics (NII)

Note that the contents and opinions expressed in this presentation reflect the views of the presenter, and do not represent the official views of the affiliated institution. 1



What is "Plasma & Fusion Cloud"?

- Interdisciplinary Use of Fusion Data, Devices, and Computers**
- "Place of Practice" for Data-driven Sciences and "Fusion Industries"
 - "Place of Education" for Expanding the Fusion Research Community





Objectives of “Plasma & Fusion Cloud”

- “Plasma & Fusion Cloud” is a new concept to realize an interdisciplinary data analysis and computation platform for both fusion experiments and numerical computations, across multidisciplinary sciences including AI/ML data-driven sciences, industries and even citizens.
- AI/ML requires huge datasets near HPC computing resources.

Education and Human/Industry Development

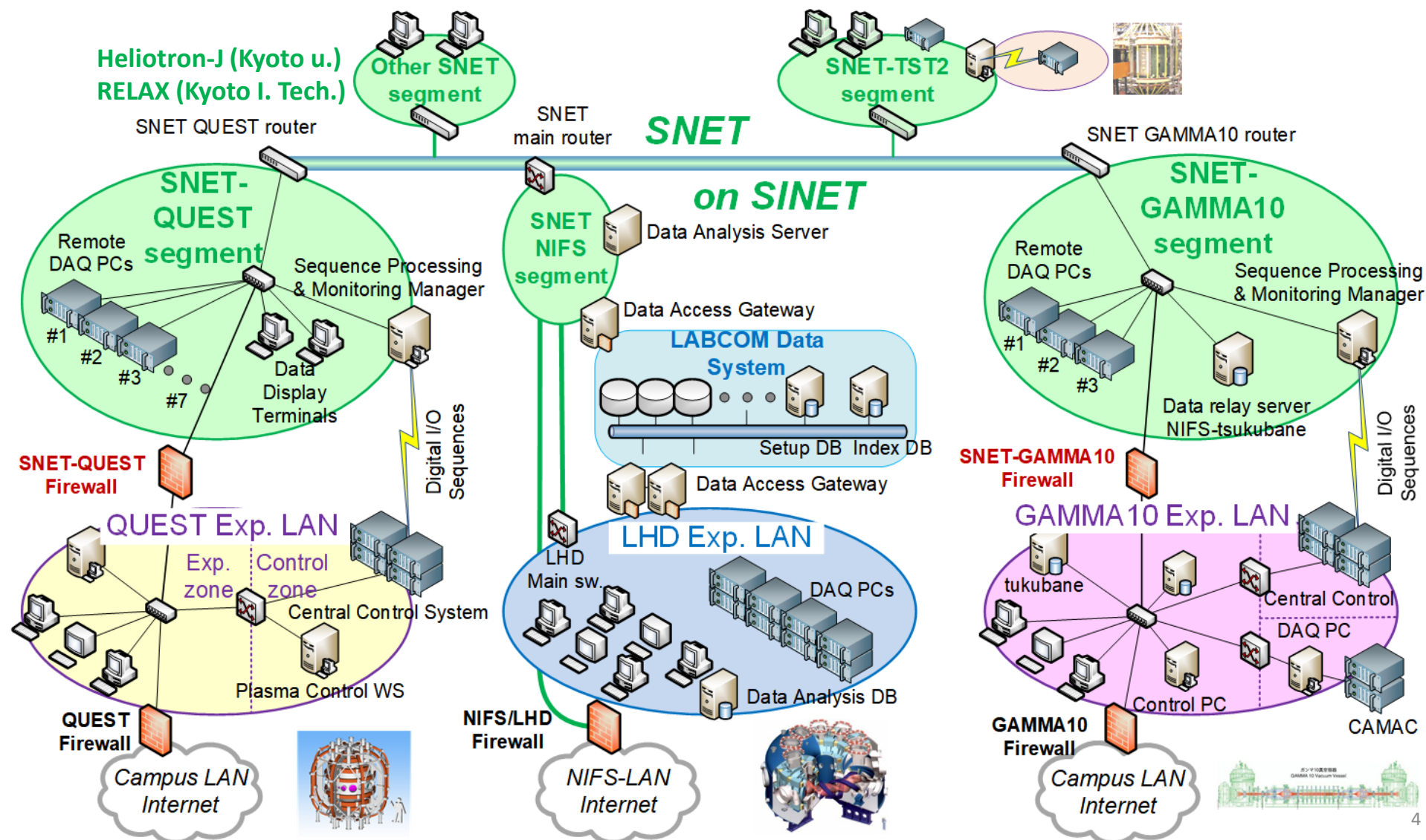
- “P&FC” is expected to be a farm place of fusion science and development in which plenty of data and computing resources are available to students and newcomers from anywhere.

Open Science & Research Data Management

- Toward “Open Science”, fusion research data should also follow the **FAIR** (Findable, Accessible, Interoperable and Reusable) data principles.
- Securing the **data traceability** becomes more important in terms of RDM.
- Assigning **global persistent identifiers (PIDs)**, such as DOI, enables the **data citation** and its counting for research achievement

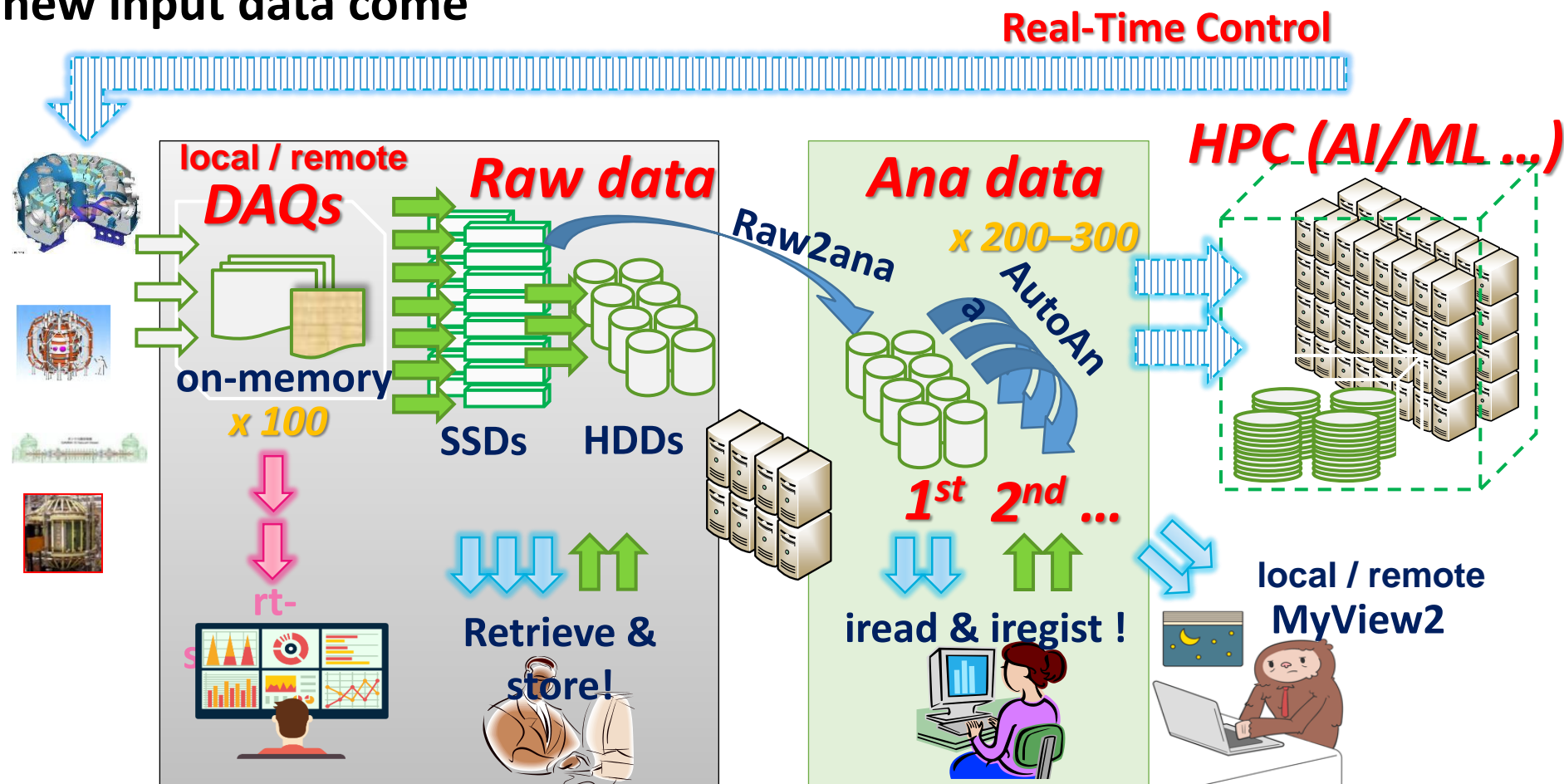
Fusion Virtual Laboratory via SNET Layer-2 VPN

SNET and the private 'Storage Area Network' are isolated from Internet, only for remote collaborations in fusion research.



Workflow chains for data processing automation

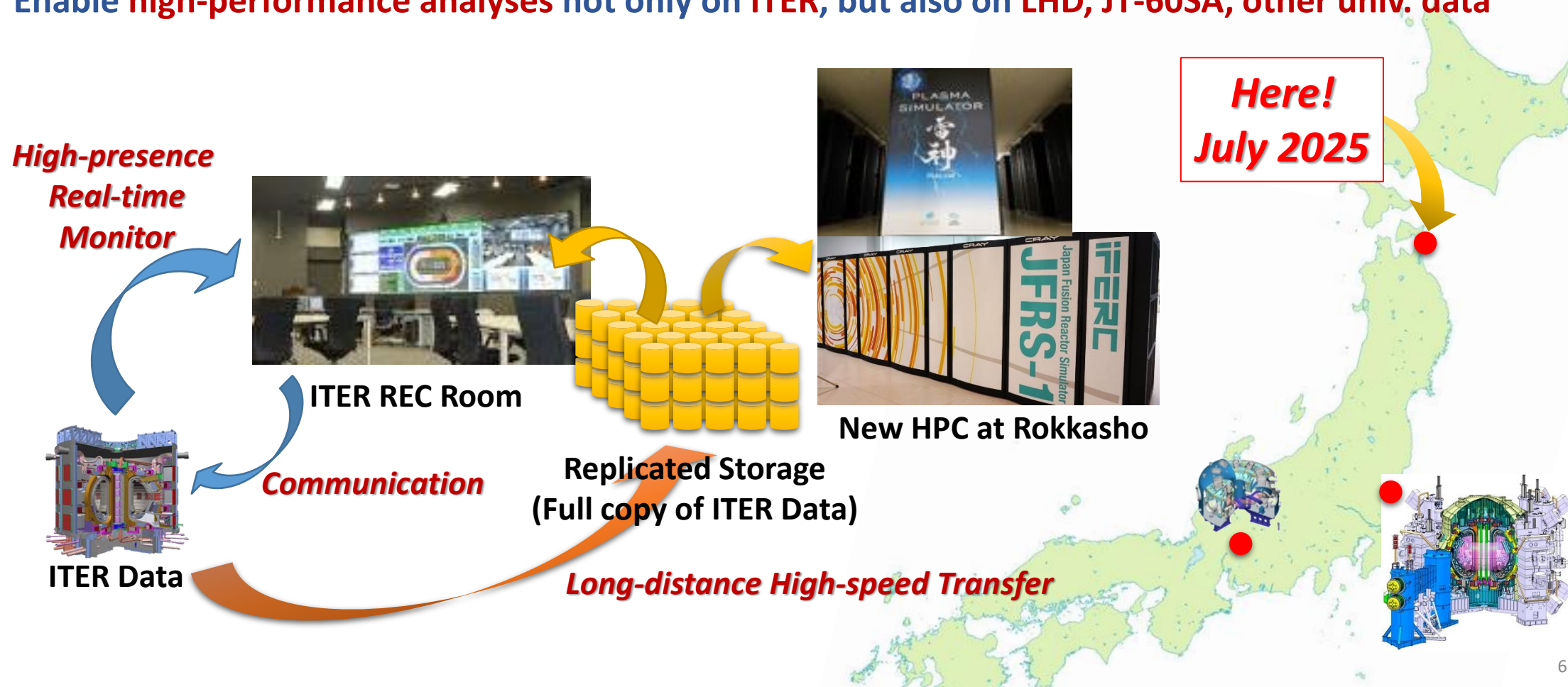
- More than 200 pre-registered programs will be triggered automatically every time when new input data come



- Real-time feedback control has been tried in 2024 LHD campaign using HPC “Raijin” cf. Morishita Y. et al., *Scientific Reports* (2024).

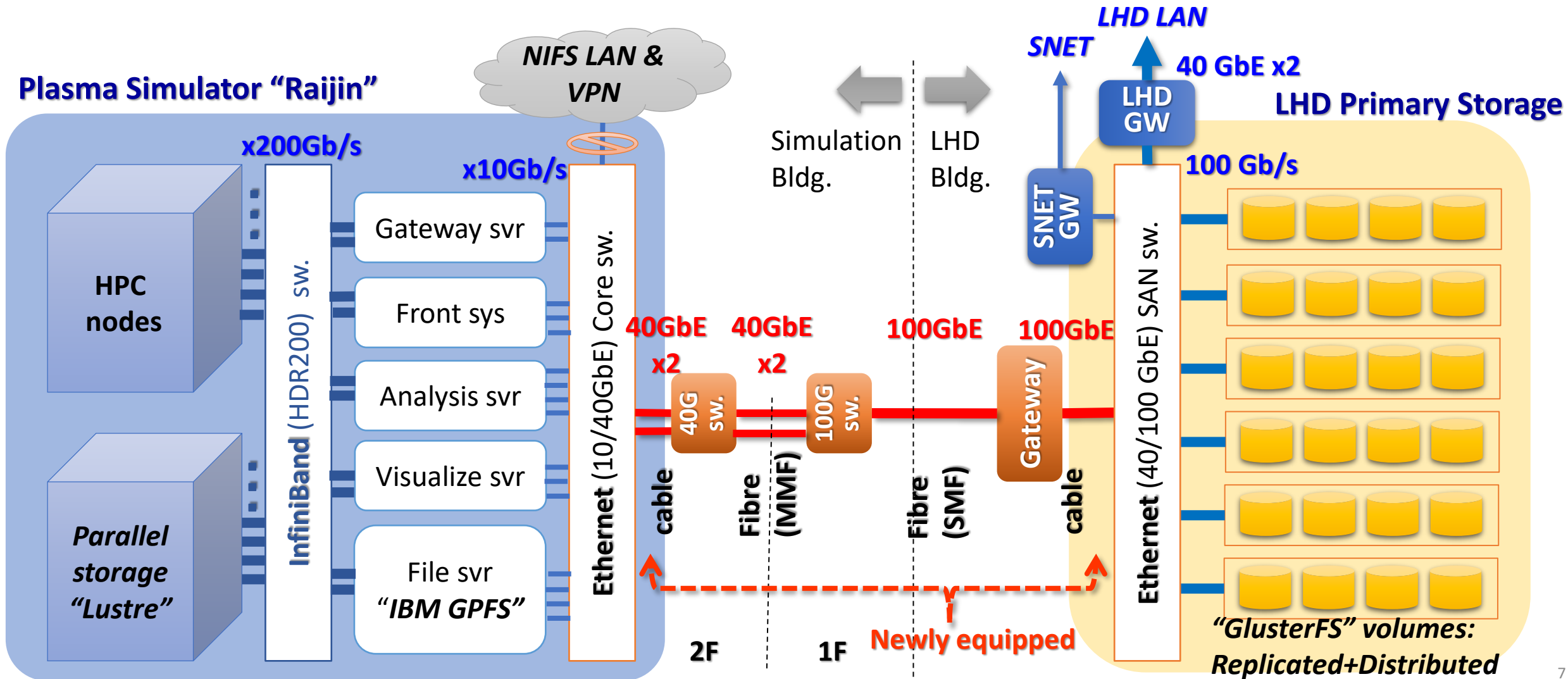
HPC as a Data Computing Platform at Rokkasho

- NIFS & QST plan to jointly procure the next-generation HPC at Rokkasho, Japan.
 1. ITER REC room will provide **high presence** as if people were on-site control room
 2. ITER data will be fully replicated to the REC storage almost in real-time
 3. Enable **high-performance analyses** not only on **ITER**, but also on **LHD, JT-60SA, other univ. data**



Direct Link between HPC and LHD Primary Storage

- Evaluate effective readout throughput from LHD data storage to HPC node's
- Also testing IMAS including UDA on HPC is now ongoing ...

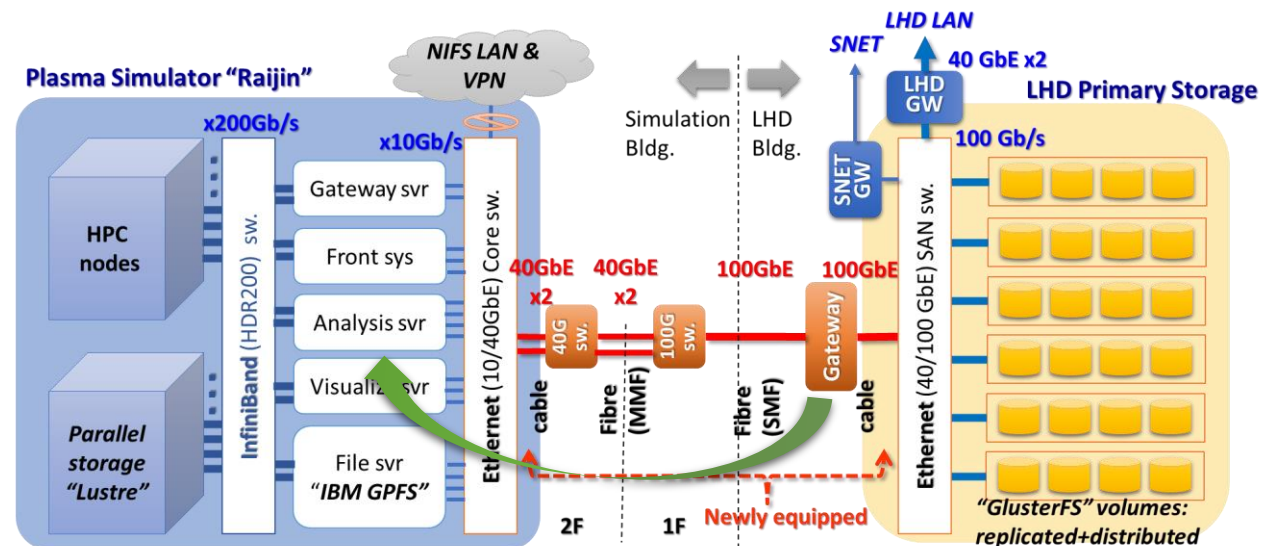


Throughputs between HPC “Raijin” and LHD Primary Storage

Effective throughputs from LHD data storage to HPC node,

✓ via 2 x 10 GbE LACP. Single TCP socket stream. Averaged value over 5 times

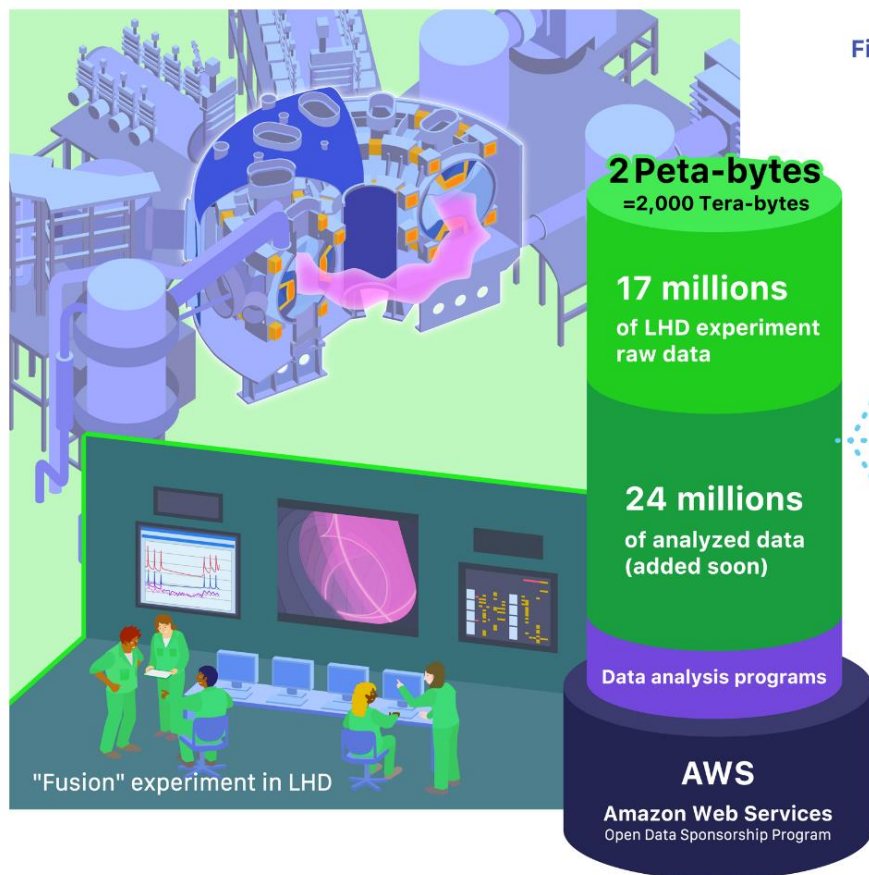
Diag	Raw size /GB	Arc size /GB	Single Thread /s	Speed (raw) MB/s	Speed (arc) MB/s	Multi Thread /s	Speed (raw) MB/s	Speed (arc) MB/s
ECE-UFAST (7)	16.0	13.5	119.57	134	112	39.68	403	340
FIR-BEAT (12)	4.8	2.68	29.74	161	90	9.34	513	286
<i>Curl/ftp 1st</i>	--	12.5	18.84	--	700	--		
<i>Curl/ftp 2nd ...</i>	--	12.5	11.51	--	1122	--		





Press Release on June 14th

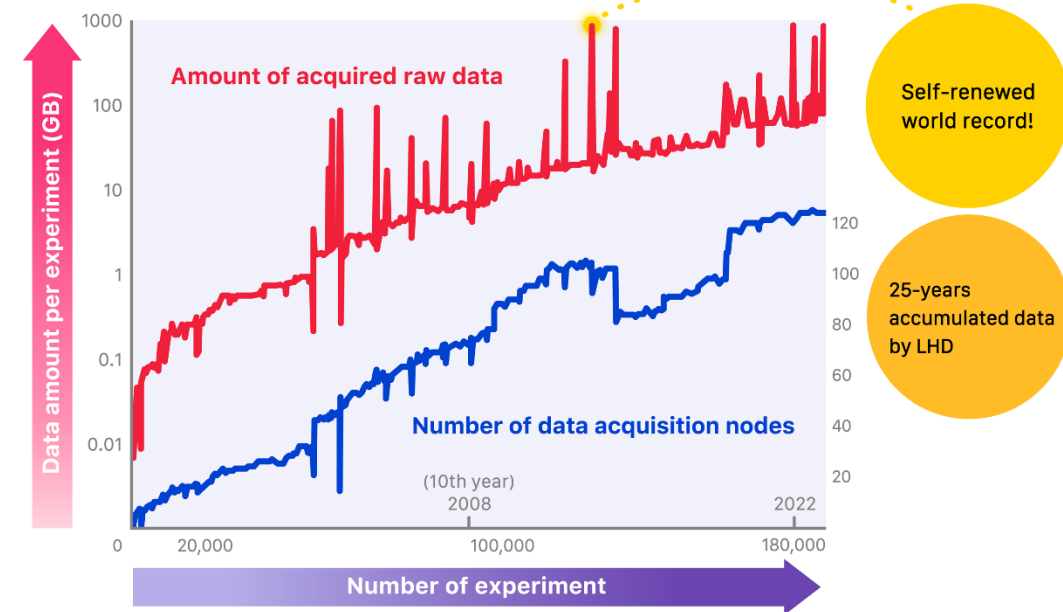
- 25 years of LHD's massive fusion experiment data completely open on the Amazon Web Services (AWS) S3 cloud storage, to be available to anyone
 - ✓ With a support of AWS Open Data Sponsorship Program (max. 2 PB on S3 freely)
 - ✓ No embargo time → A major step toward making fusion research an "Open Science"



First open data of "Fusion" experiment



0.92 PB/shot acquired in a plasma pulse



Total amount of 25 years accumulated diagnostic data is 1.5 PB in zip archived size.

DOI registration of 40M fusion physics data ... another challenge



- DOI registration has been started since August 2023 for LHD's fusion diagnostic data → FAIR Principles
 - ✓ 1.22 million DOIs have been issued on diagnostic data objects by November 2023, such as <https://doi.org/10.57451/lhd.fpellet.182000.1>
 - ✓ Each DOI access is redirected to the corresponding landing page which is dynamically generated by the PHP script. →
 - ✓ First challenge to give DOIs on such enormous number of fusion experiment data.
 - ✓ 40 million raw & analyzed data need at least several years to complete registrations with a specific cost... Big problem.
 - ✓ Top research data DOIs: Geoscience Australia (approx. 7 million DOIs), CERN (approx. 6.7 million), Interdisciplinary Earth Data Alliance (approx. 5 million) → We can/should do it, together!

Fusion Diagnostics Data Repository
LHD FPellet #123456.1

Data description
Fueling Pellet Injectors (FPellet): Inject hydrogen isotope pellets into core plasma to control plasma density. The in-situ pipe gun type 20-barrel injector and the screw extruder type repetitive pellet injector are integrated. [More](#)

Creator(s)
SAKAMOTO Ryushi^{1,2,3}, MOTOJIMA Gen^{1,2,3}, YAMADA Hiroshi^{1,2,3}
National Institute for Fusion Science (NIFS)

Acquisition status
Completed with no errors Nakanishi H. (中西 秀哉) (naka...)

Primary info.

Shot	123456
SubShot	1
Management	10.0.0
AcquisitionDate	Thu, 05 Dec 2013 16:43:22 +0900
ArchiveDate	Thu, 05 Dec 2013 16:43:22 +0900
ModuleGroup	WE7000
Modules	16
Channels	15
AcquiredChannels	15
Comment	
Rights	Copyright by National Institute for Fusion Science (NIFS).
Access Rights	Rights and Rules of LHD data
APC	Not charged
Publisher	National Institute for Fusion Science (NIFS)
Identifier	https://doi.org/10.57451/lhd.fpellet.123456.1
Registration	JaLC DOI
Relation	Development of Advanced Pellet Injector Systems for Plasma Fueling (2009) Twenty barrel in situ pipe gun type solid hydrogen pellet injector for the Large Helical Device (2013)
Funding Reference	Large Helical Device (LHD) project, NIFS
Resource Type	Dataset
Version	1.0
File	Name : FPellet-123456-1.zip Size : 47750191 Type : zip Download

Channel parameters ch# < > 1 2 3 4 5 6 7 8 9 10 11 12 13 >

Name	Value	Type
Shot	123456	INT
SubShot	1	INT
ChannelNumber	8	INT
ModuleGroup	WE7000	STRING

DOI landing page:
[10.57451/lhd.fpellet.123456.1](https://doi.org/10.57451/lhd.fpellet.123456.1)



Summary

“Plasma and Fusion Cloud” data ecosystem has been technically verified on some issues.

Enormous amount of diagnostic data inevitably require a HPC platform nearby, not only for LHD/SNET but also for ITER, JT-60SA, and other experimental devices.

To make the **“cloud” data ecosystem** may help to exchange the interdisciplinary intelligence not only between experiments and theoretical model calculations, but also with other fields. Cooperation with other community are very essential for **“Open Science”**.

1. **Performance evaluation have been made using the HPC supercomputer "Raijin" and the LHD primary data storage.**
 - ✓ Directly connected by 100 Gbps Ethernet link.
 - ✓ Throughput improvement is still ongoing. Multiple parallel streams might improve the speed.
2. **1.22 million DOI issued for each diagnostic data object**
 - ✓ LHD have roughly 40 million data objects by “data-name × shot-no.” basis
 - ✓ Registration fee and the number of years to complete registrations are the problems!
 - ✓ Other fields registering many data DOIs have long history of international data exchanges, cf. geo-science.
3. **Open Science – How far as open as possible and as closed as necessary?**
 - ✓ LHD has decided open everything: data, codes, proposals, summaries, logs, etc.
 - ✓ Cloud technology and its storage is a powerful supporter for scientific Open Data.
 - ✓ Data license agreements are needed with other SNET institutions, e.g., consensus for common license?
4. **Standardization in fusion community**
 - ✓ Standardization for data platform; UDA and IMAS compatibility checking and implementation are ongoing.

Let's Collaborate on “Cloud”!



Let's Collaborate on "Cloud"!