Open access project at KEK

Konomi Omori , KEK CRC

FJPPL – France-Japan workshop on computing technologies
18 - 19 February 2025, Lyon, France

Motivation

- KEK is a research institute that collaborates with partners worldwide on joint research projects.
- KEK promotes open science and is working to achieve open access (OA) to research data.
- To support OA, KEK is planning to introduce a research data management system to store and publish academic publications and related materials.
  - The system can connect with other institutes around the world through federated authentication, that brings various benefits.
  - For example, sharing research may lead to new studies by combining different findings.
- KEK is highly motivated to join an authentication federation, which is essential for achieving OA.

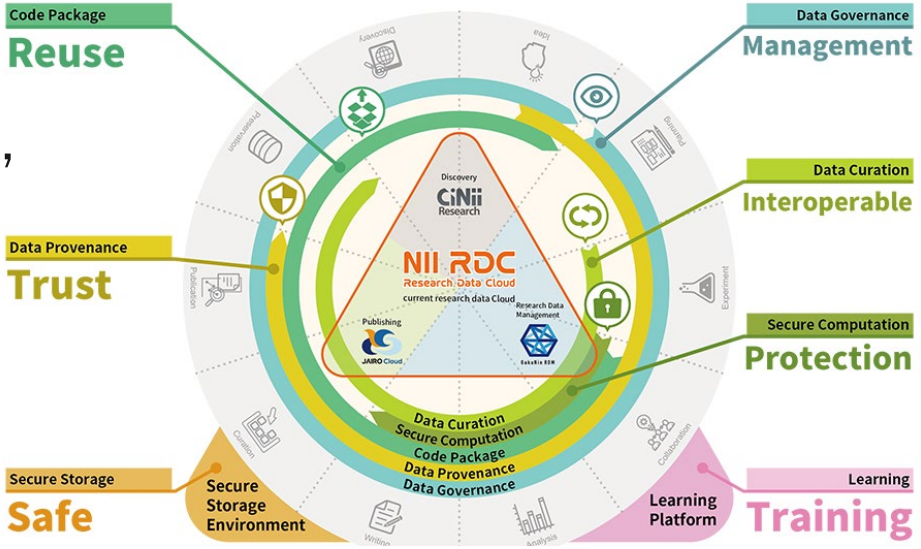
Background - Mandatory Immediate OA

- In Japan, from fiscal year 2025, researchers who receive public funding (competitive research grants) will have to make their academic papers and supporting data immediately open access (OA).
- The government is promoting an OA acceleration program to support universities and research institutes with the costs needed to create an environment for immediate OA. KEK was selected for this program.
- With this support, the OA project was launched at KEK Computing Research Center (KEK-CRC).
- As part of this project, we are trying to introduce the **Research Data Cloud** provided by NII* as a research data management system.

* National Institute of Informatics (Japan)

KEK's Plan to Use NII RDC

- The NII Research Data Cloud (**NII RDC**) is an information platform for managing, publishing, and utilizing research data efficiently.
- It consists of three platforms: research data management platform (GakuNin RDM), publishing platform (JAIRO Cloud), and discovery platform (CiNii Research).
  - **GakuNin RDM** (research data management) is a cloud platform for managing and sharing research data.
    ⇒KEK-CRC is trying to implement this system, and this is a main topic of this presentation.
  - **JAIRO Cloud** is a repository system for publishing research data and papers.
    ⇒KEK plans to migrate its current repository to JAIRO Cloud.
  - **CiNii Research** is an academic search service linking papers to researchers.
    ⇒KEK already uses it.



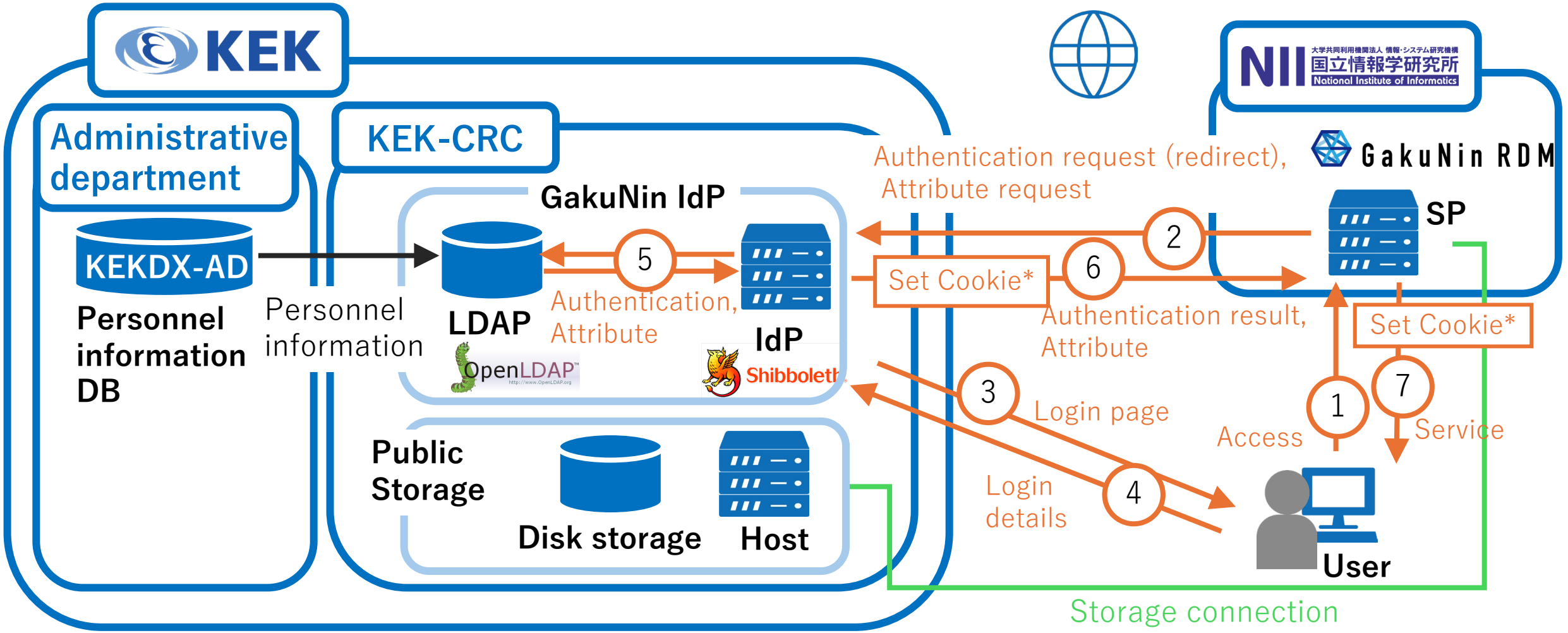
What is GakuNin?

- **GakuNin** is the Academic Access Management Federation in Japan and is based on **Shibboleth** technology.
- In GakuNin, there are **Service Providers** (**SPs**) that offer services, and each institute has its own **Identity Provider** (**IdP**) to authenticate users.
  - SPs do not authenticate users directly but rely on IdPs. They also request user attributes from IdPs as needed.
  - By trusting the policies defined by the federation, institutes can enable federated **authentication with each other**.
- **To join GakuNin**, we **must have an IdP server** for authentication.
  - An IdP server needs a database that includes personal data for authentication and attributes.
- **GakuNin can connect with eduGAIN** to share services with the European Academic Access Management Federations.

To Introduce GakuNin RDM

- GakuNin RDM is one of GakuNin's SPs.
- We need to **set up a new IdP** and join GakuNin to use GakuNin RDM.
- We will **set up an LDAP server** as a database to store user information for the IdP.
- A **public storage system** is also needed to store research data in GakuNin RDM.
- Currently, KEK-CRC is working on **joining GakuNin** and **building public storage** as part of the OA project.

Functional Relationship Map

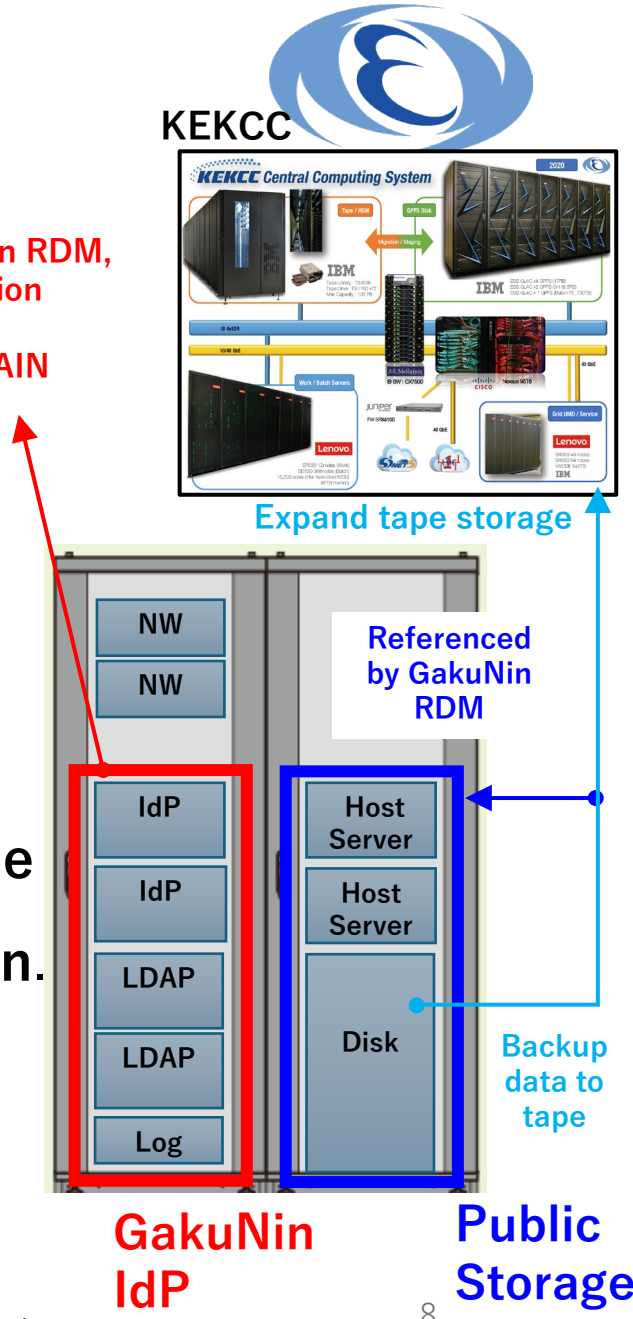


*Cookie: Store the authentication result by the IdP

System Overview

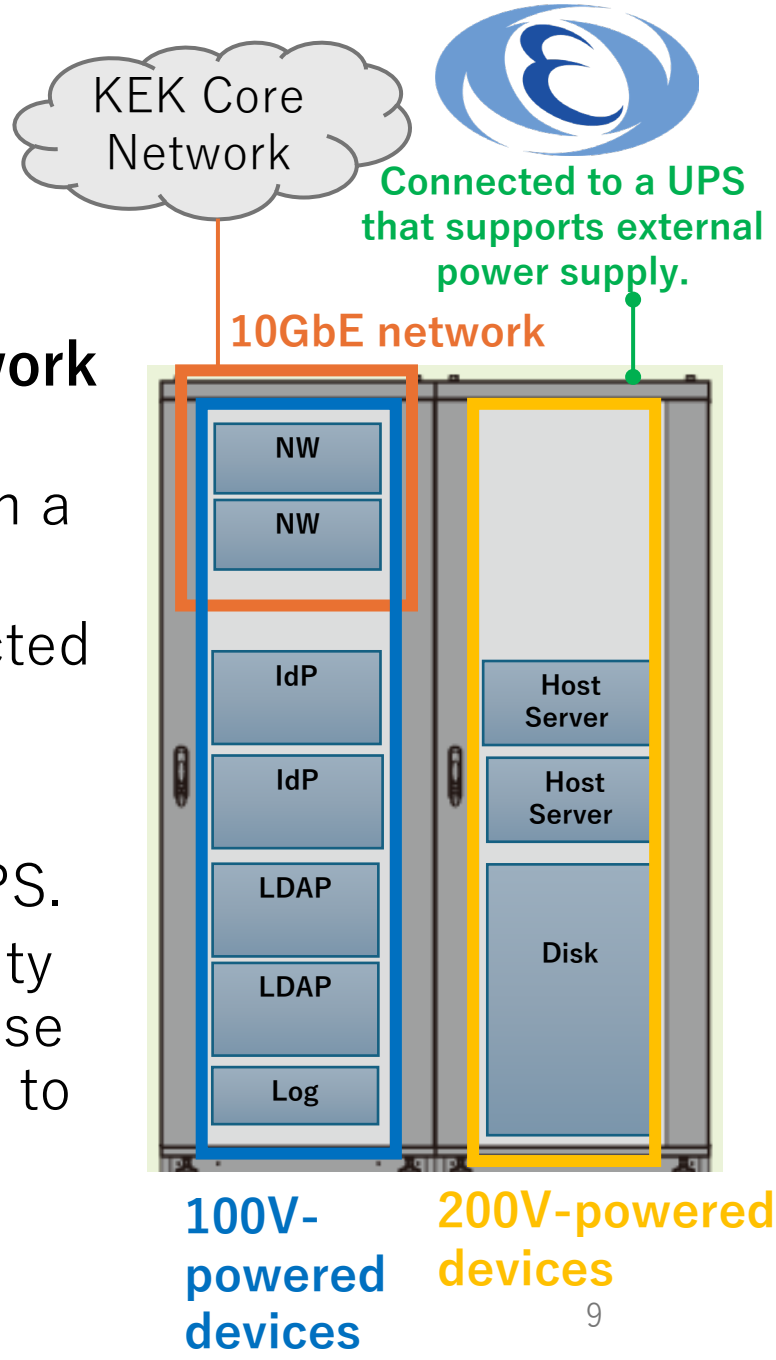
- **Strategy**
  - This year, **~430k euros** is available.
  - **Hardware procurement is in progress**.
  - The service is expected to **start in 2025** using this equipment.
  - The system is expected to operate **until 2028**.
  - After that, it is planned to be integrated **into the next KEKCC** to reduce running costs.
- **Components**
  - **GakuNin IdP**
    - We are preparing **two IdP servers, two LDAP servers,** and **one log server**.
    - The IdP and LDAP servers will have **a redundant configuration**.
  - **Public Storage**
    - We are preparing **two host servers and 2PB of disk storage**.
    - We plan to deploy **S3-compatible software** (like MinIO*). We are considering which software to use.
    - We are adding **2PB of tape storage** to KEKCC **for backup**.

*https://min.io/



Infrastructure Setup

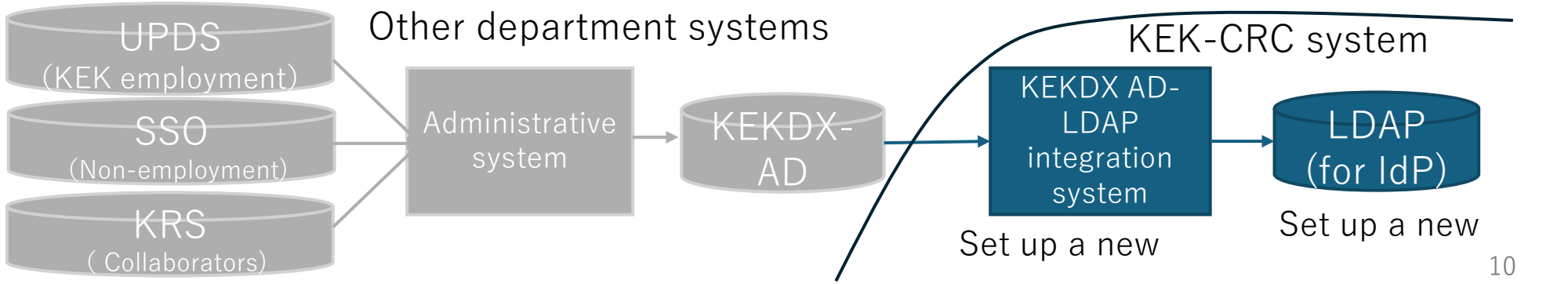
- **Network**
  - The **OA network** is connected to the **KEK core network** via **10GbE** to support large-scale data transfers.
  - Two **network switches** will be provided and set up in a **redundant configuration**.
  - The IdP and public storage are planned to be connected to the DMZ.
- **Power**
  - This system must operate without interruption by UPS.
  - KEK has **scheduled annual power outages** for facility maintenance. To keep the system running during these outages, an **external power supply route** is planned to be set up.
  - **New distribution boards and power cables** will be installed for this purpose.



LDAP Server and User Information

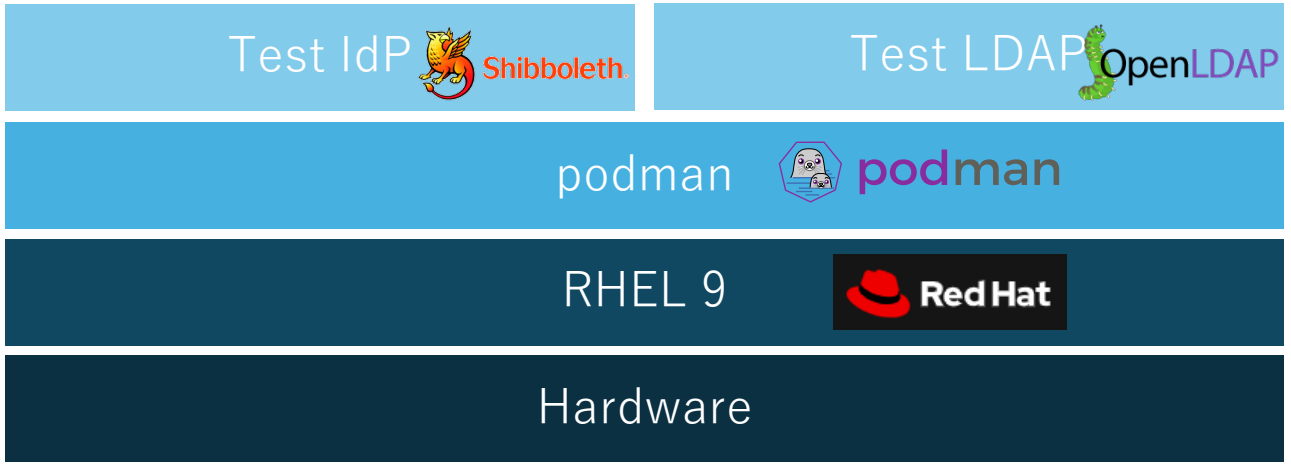
- The LDAP server used by the IdP needs KEK staff information with specific LDAP attributes.
  - This information will come from the existing Active Directory (KEKDX-AD).
  - KEKDX-AD is managed by the administrative department and stores personnel information that meets IAL2 (Identity Assurance Level*).
  - Until now, KEK-CRC has not been integrated with administrative systems. This will be the first challenge that LDAP is connected to an administrative system.
  - This integration will allow KEK-CRC to refer ID information that meets IAL2 standards.
- We need to prepare a system to synchronize data from KEKDX-AD to LDAP (for IdP), and we are currently considering it.

* https://csrc.nist.gov/glossary/term/identity_assurance_level



Test Setup of IdP Server

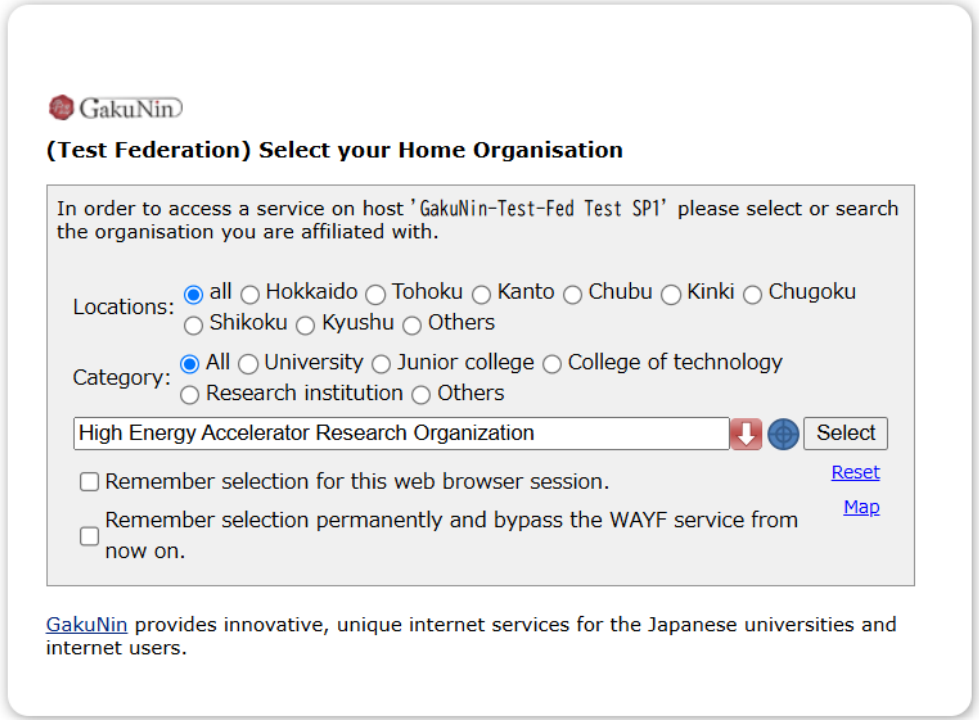
- We set up a test IdP server.
- It was built with Shibboleth.
- The IdP server and LDAP server were built using container technology.
  - Since Docker is not supported on RedHat 8 and later, we used podman.
  - Using podman allows easy replication and migration.
  - This system can be quickly started with the `podman-compose` command.



Test server configuration diagram

Joining a Test Federation

- GakuNin has both a test and an operational federation.
- We succeeded in joining the test federation using our test IdP server.
- We are going to join the operational federation by the end of 2025.



KEK IdP selectable on the test SP's IdP selection screen

Personnel information successfully recognized within a test SP

Future Plans

- We are going to complete the hardware procurement and power setup by March 2025.
- By March, we will set up the production IdP server, join the GakuNin operational federation, and enable initial use by some users.
- KEK's IdP server aims to connect to eduGAIN through GakuNin, which could enable collaboration with European Academic Access Management Federations.
- This system will be part of the next KEKCC, planned for 2028.
- NII* is considering adding IAL2 support to GakuNin accounts. Grid access requires IAL2 verification, so if GakuNin supports IAL2, it could be used for Grid authentication.
- In the future, we aim to replace Grid authentication with GakuNin.

* National Institute of Informatics (Japan)
NII developed GakuNin.