

International Workshop on Sharing, Citation and Publication of Scientific Data across Disciplines 2017
Joint Support-Center for Data Science Research (DS), Tokyo

Development of Research Data Management Service for Open Science in Japan

Research Center for Open Science and Data Platform (RCOS),
National Institute of Informatics (NII)

Yusuke Komiyama

komiyama@nii.ac.jp

5, December 2017

Open science and research integrity in Japan

- Domestic situation
 - First, in Japan, all universities and research institutions need to preserve research data from each laboratory for ten years because it saves a **research integrity**. And it needs to be the service which contributes to **open science**. However, the nationwide infrastructure of **research data management (RDM)** for general purpose does not exist in Japan yet.
 - Second, National Institute of Informatics (NII) is providing the nationwide IT infrastructure for academia such as a high speed network and federated authentication, for university now.
 - Third, NII is going to develop the RDM service and deploy to 850 universities and research institutions, and we will integrated research-data infrastructures from three fundamental platforms for research integrity and open science, it includes discovery, publication and management platforms.



Research Data Platforms for Open Science

CiNii Research

Discovery platform

- Linking Func between Article and Data
- Researcher and Research Project Identification and Management Func
- Data Exchange with International Discovery Service

GakuNin RDM

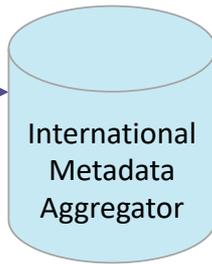
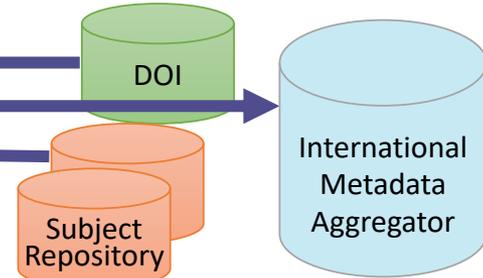


Research Data Management System



- High Speed Access using SINETS
- Data Sharing Func using Virtual NW and ID Federation
- Effective Data Storage Switcher

Discovery Service



Metadata Aggregation



Search/Find

Data User

Data Depositor

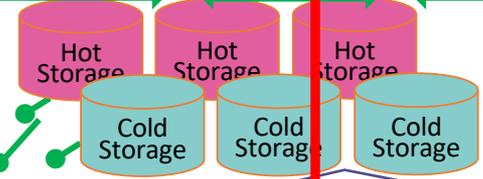
Exp/Store

Archive

Exp Data

Article

Private Shared Public



Storage Area for Long-term Preservation

Journal Article

Supplemental Data

Institutional Research Data Mng

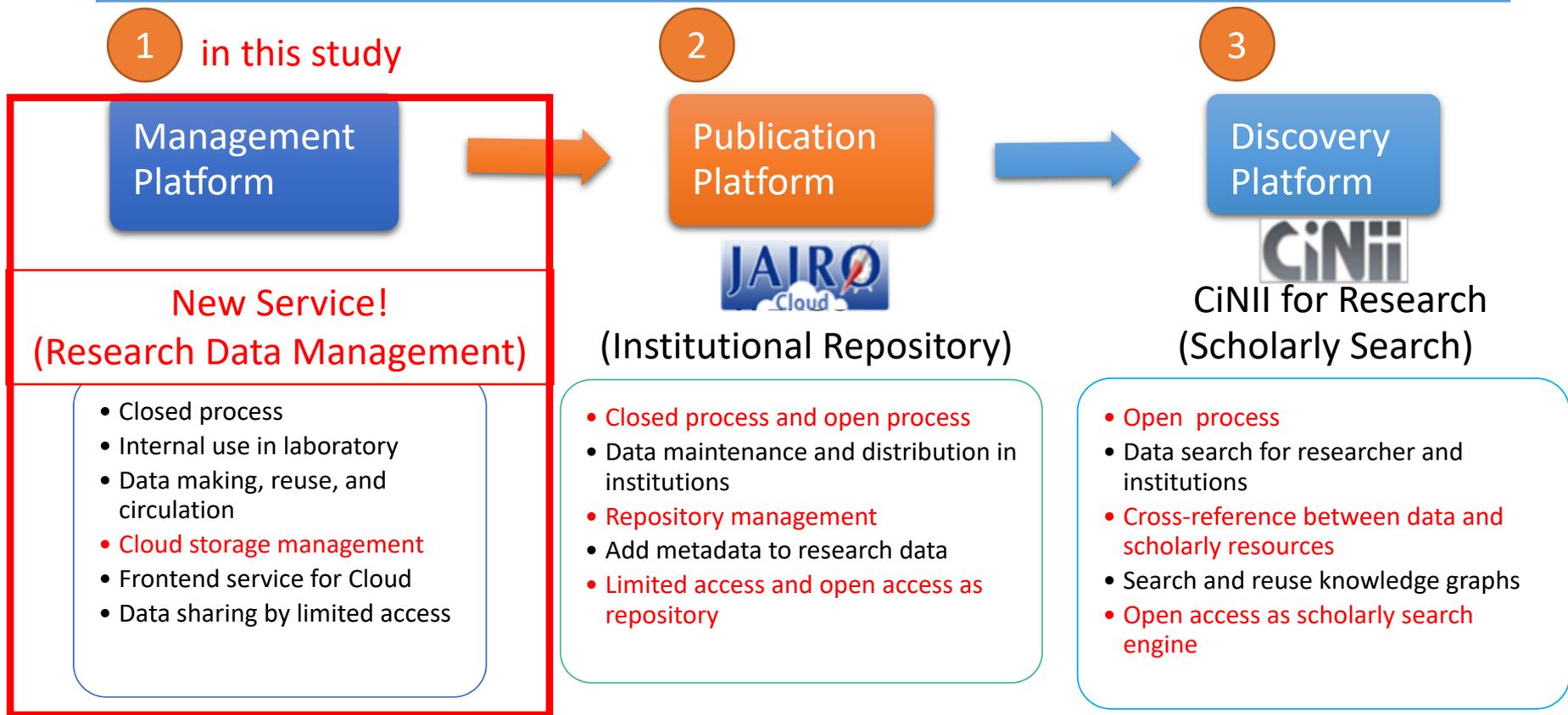
Research Data Repository

Publication platform

JAIRO Cloud by WEKO3

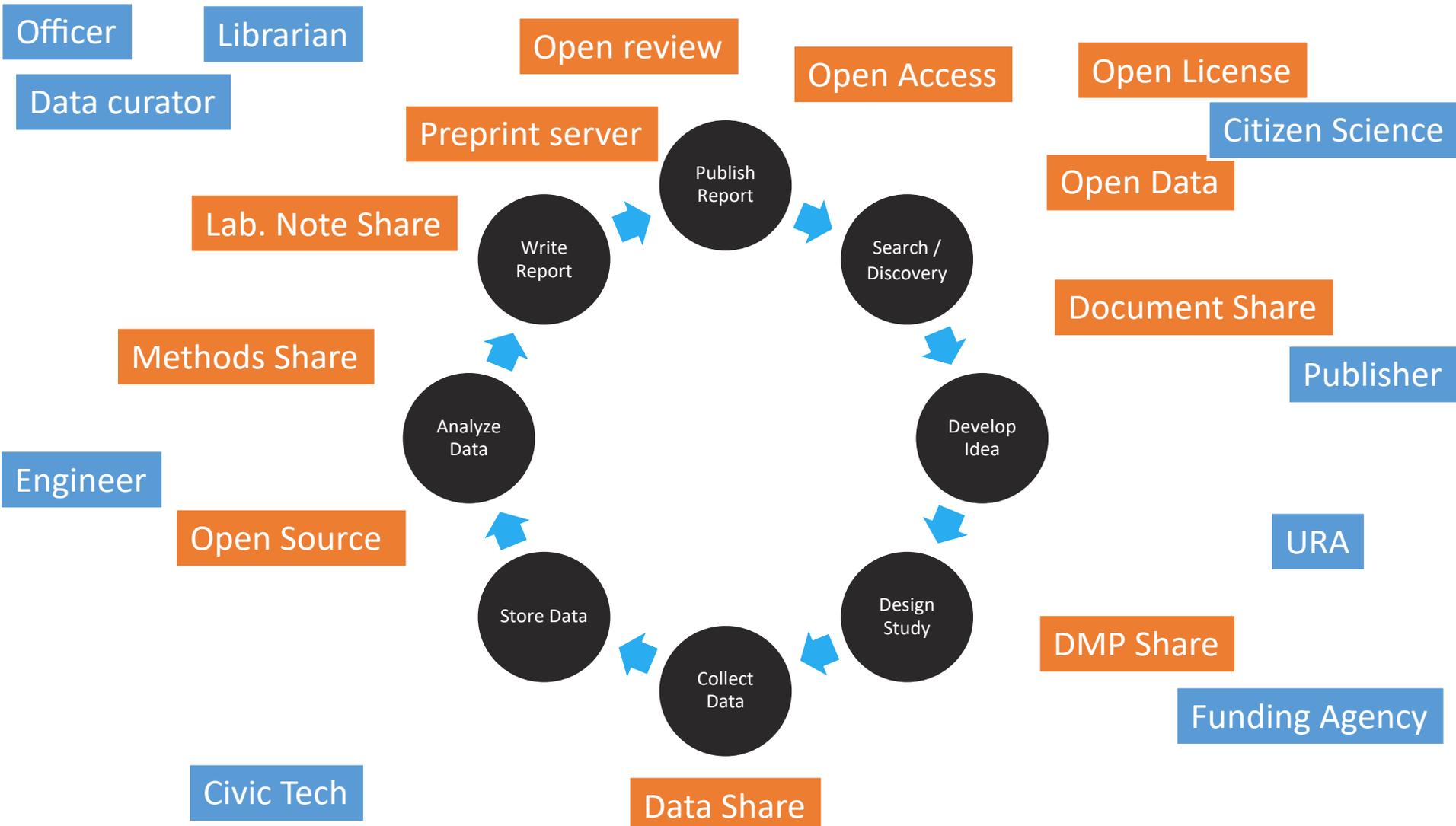
- Data oriented Self-Archiving Func
- Versioning and auto-Packaging Func
- User Dependent Personal Data Pseudonym Func

How to gather distributed research data, and make national-scale knowledge graphs for research



Summary of data flow as scholarly resource in NII research data platforms

Research cycle of open science



DMP: data management plan

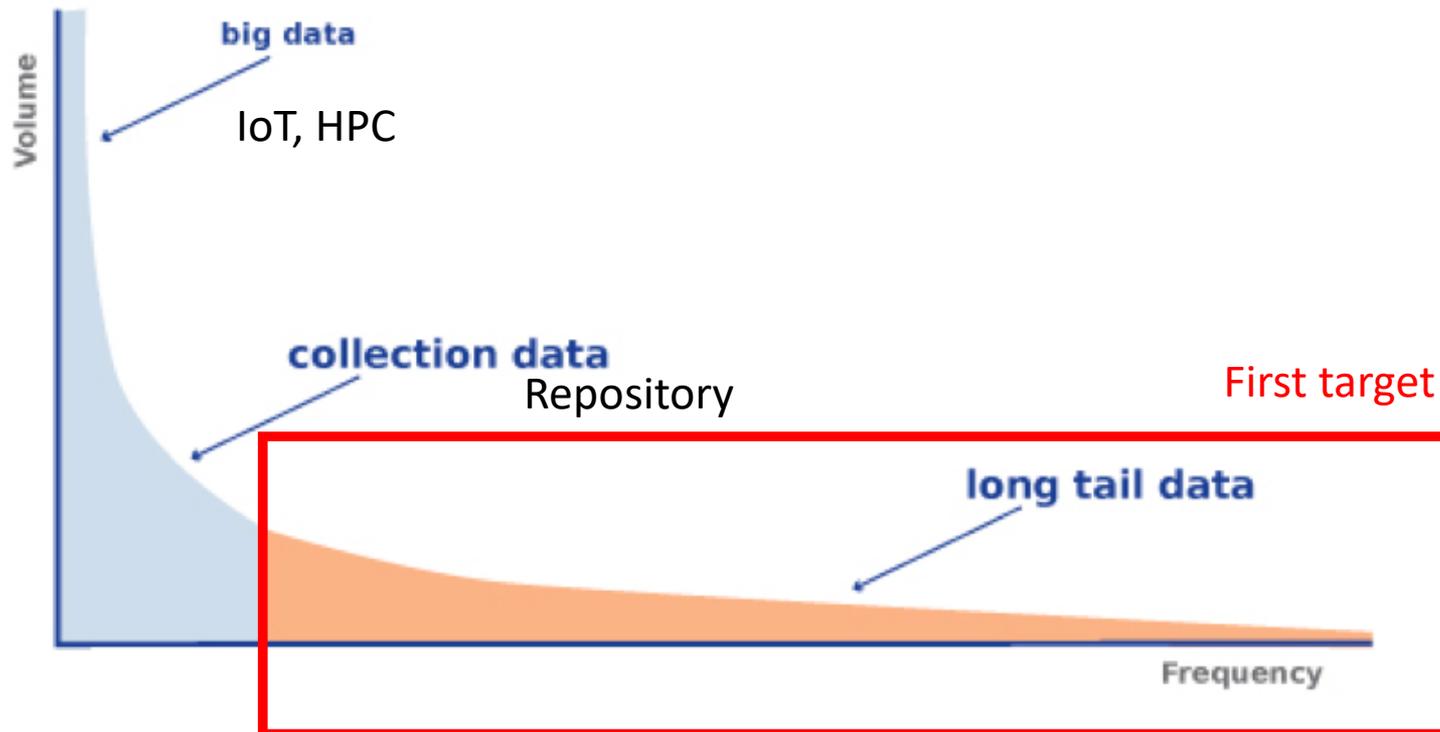
GakuNin RDM: research data management service

- NII is going to develop and provide the sustainable service for research data management (RDM) service to IT center and library in university in Japan. Purpose of this project is **research-integrity saving** and **open science promotion**.
 - We are going to mature GakuNin RDM as research data management service for Japanese academic culture and society. We cooperate closely in IT service with the university and institutions.
 - We promote the sustainable management of research data.
 - We will investigate a research tool service with GakuNin RDM

Stakeholder of GakuNin RDM

- We will discuss the collaborative possibility of GakuNin RDM service with stakeholders.
- **Stakeholder**
 - **CIO of IT center or supercomputer center in university and research institutions**
 - **Librarian in university library and research library**
 - enterprise
 - research hospital
 - funding agency
 - academic society

Long tail data and big data in research data



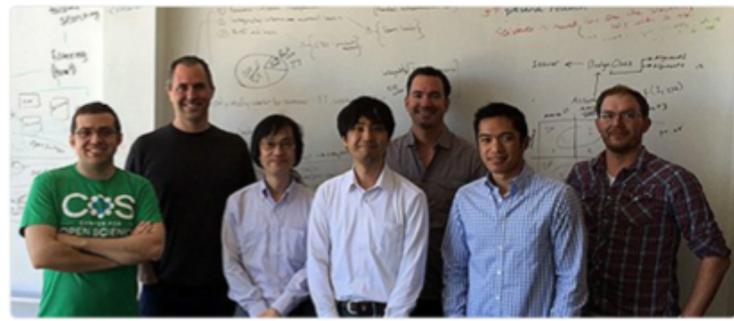
<https://www.eudat.eu/eudat-service-road-map>

Long tail data include useful knowledge for data science in academia. However, various research data is not collected, its data was lost until now. long tail data is first target for GakuNin RDM service.

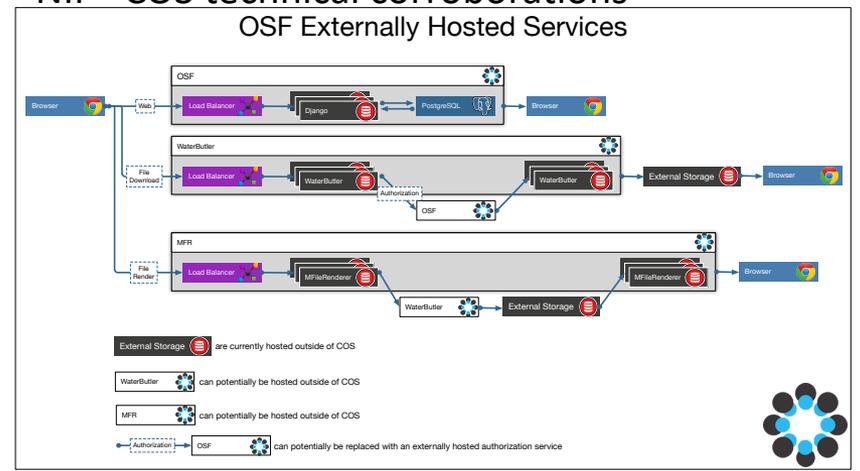
Center for Open Science (COS)

NII visited COS office last year

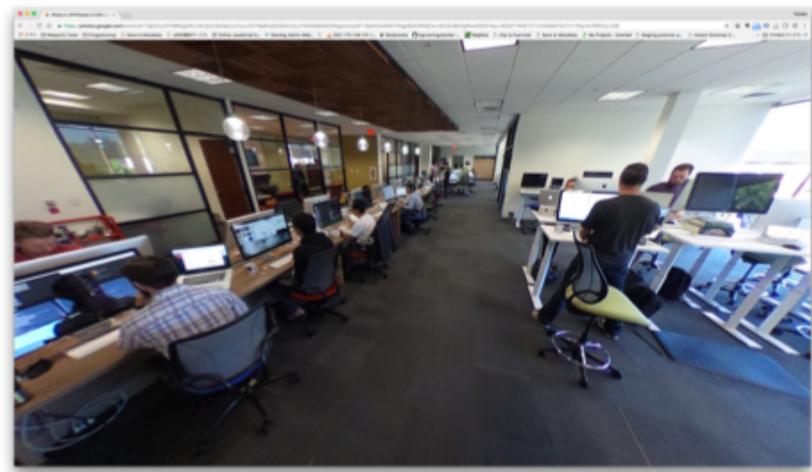
 **CenterForOpenScience** @OSFramework · Sep 14
 Great visit yesterday from @takechan2000 and @YusukeKomiya
 from NII Japan nii.ac.jp/en/ Thank you!



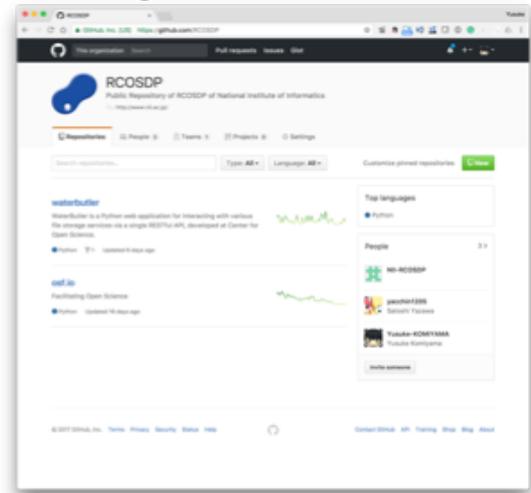
NII—COS technical corroborations



COS office in Charlottesville, VA, US

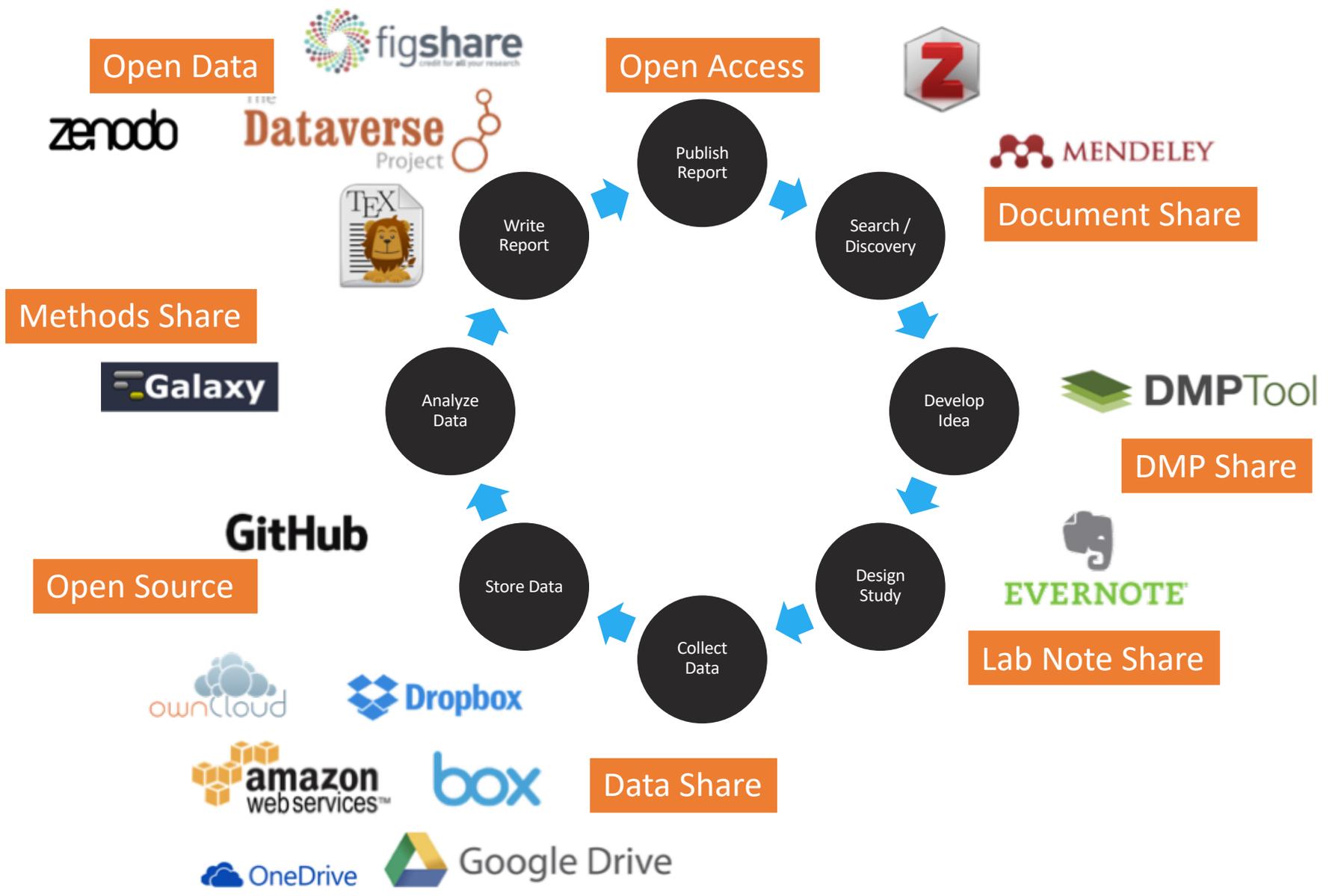


Source cord sharing with NII



<https://goo.gl/photos/WmoHMs3s7ouDbBN9>

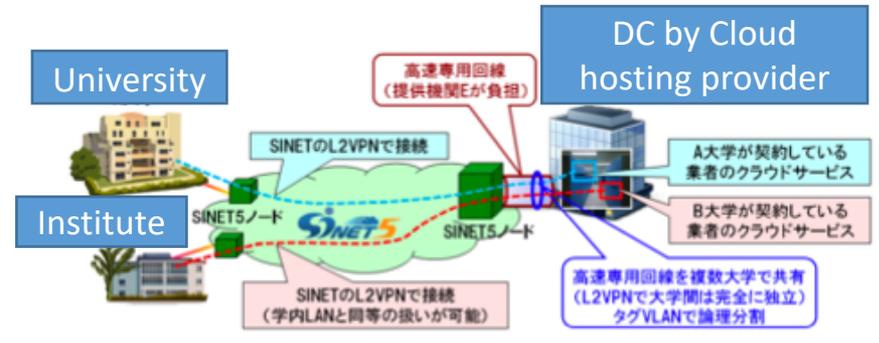
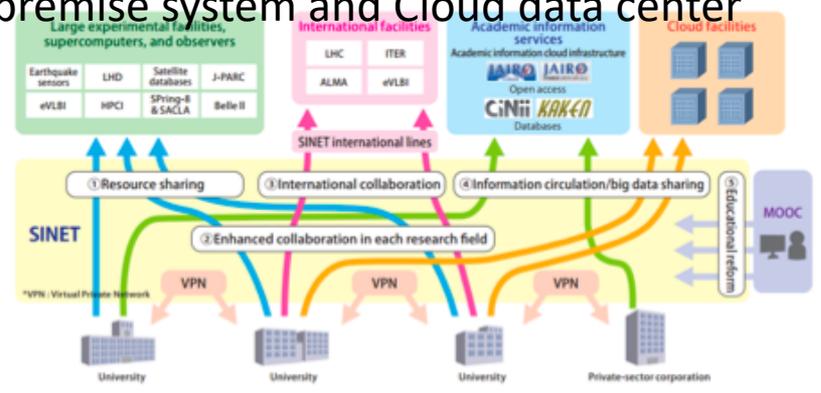
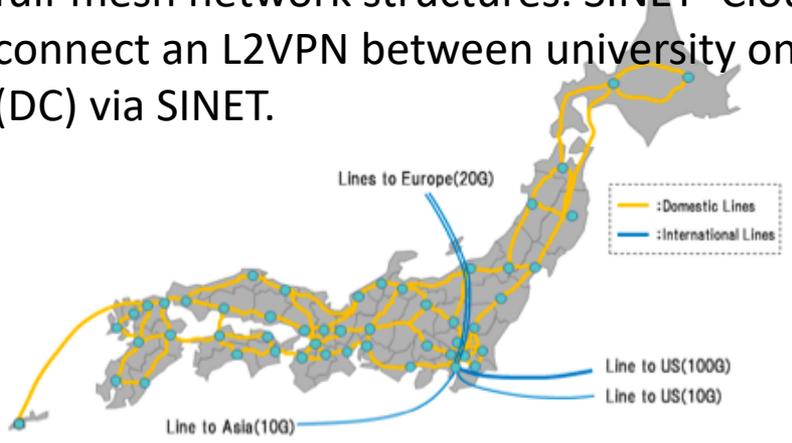
Research cycle of open science using OSF



SINET5-Cloud connection service

SINET5:

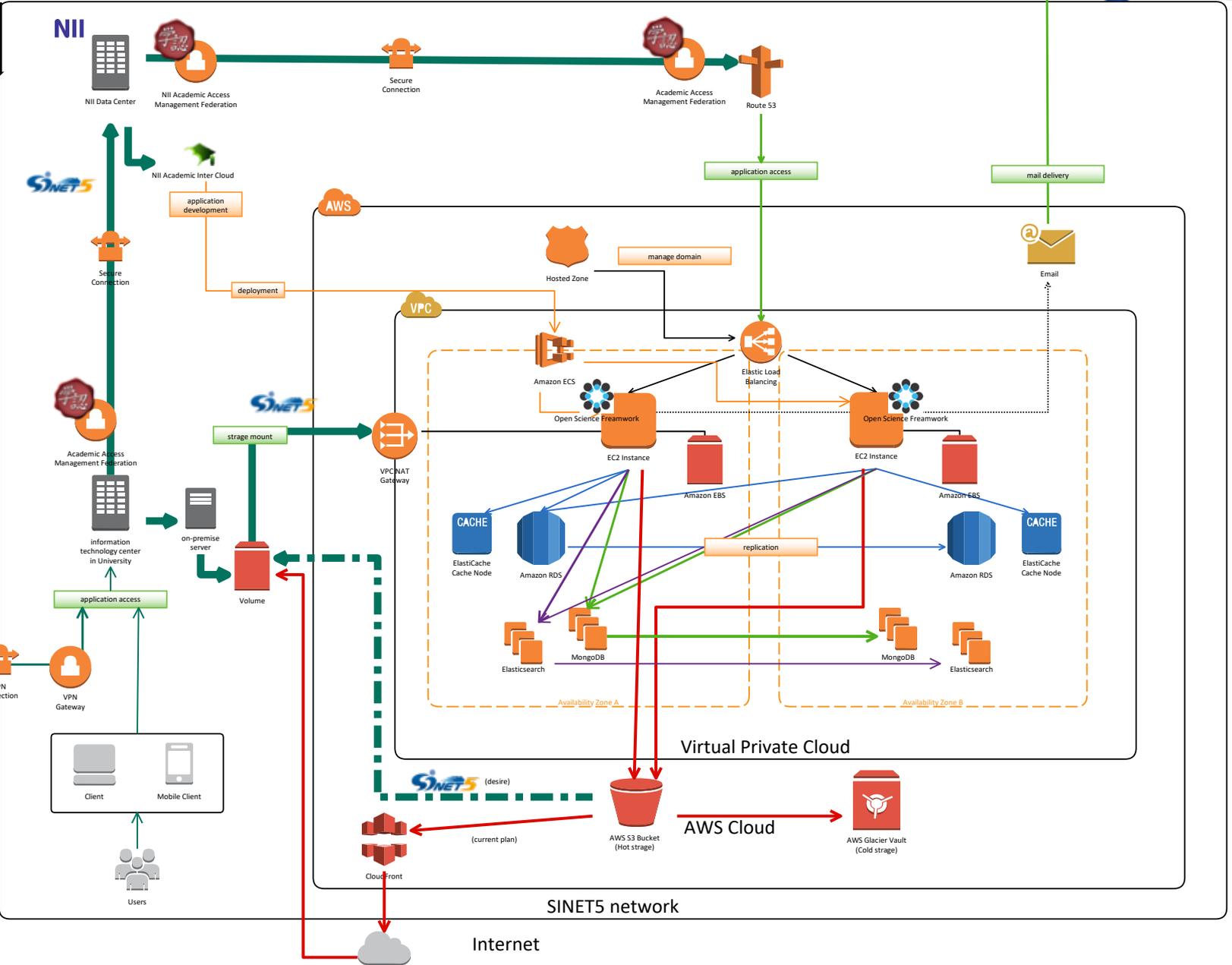
Science Information Network (SINET) is academic backbone network in Japan for more than 850 universities and research institutions. SINET5 achieved 100Gbps full-mesh network structures. SINET-Cloud connection service provides to connect an L2VPN between university on-premise system and Cloud data center (DC) via SINET.



19 Japanese Cloud providers

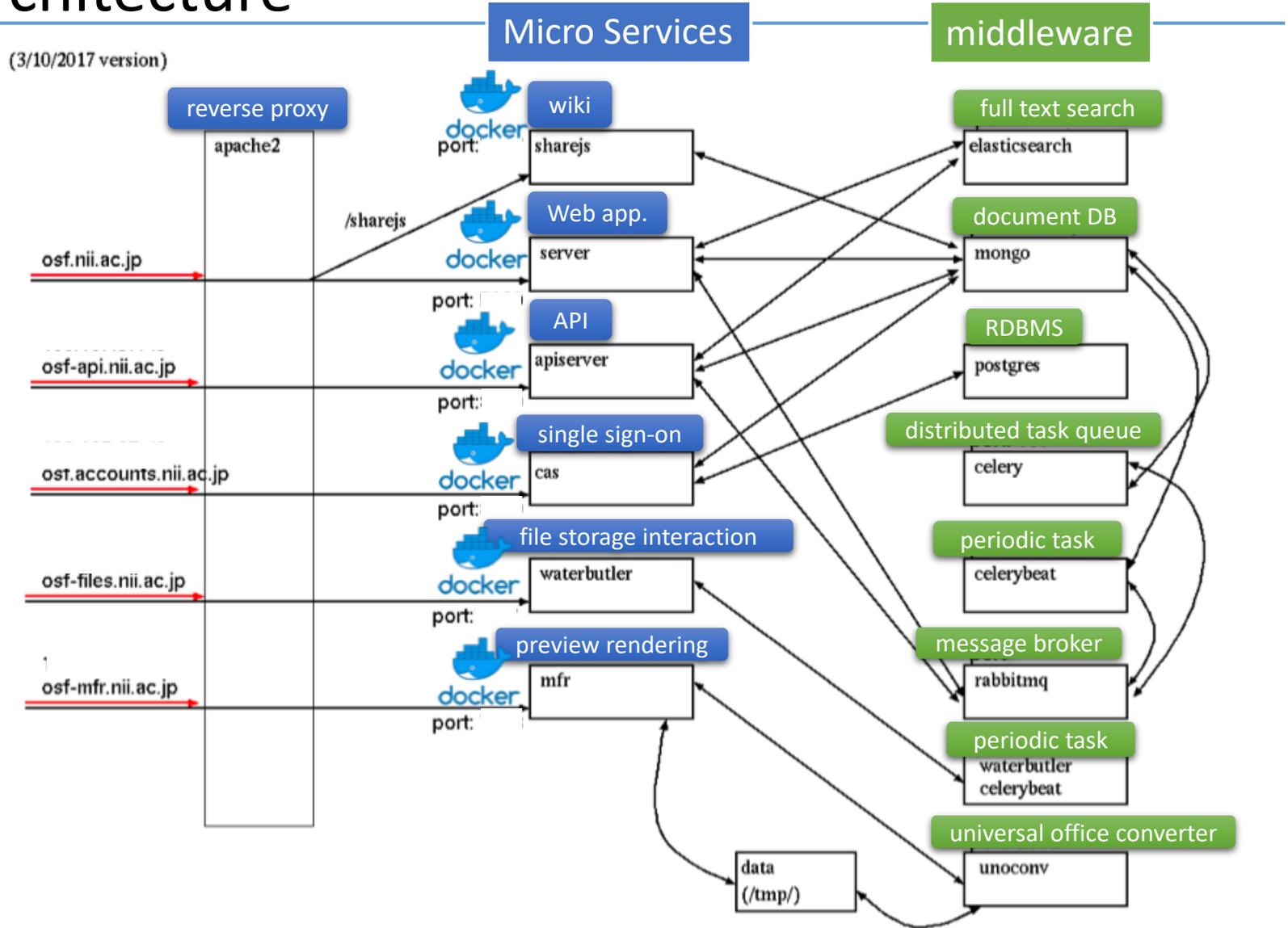
Microsoft

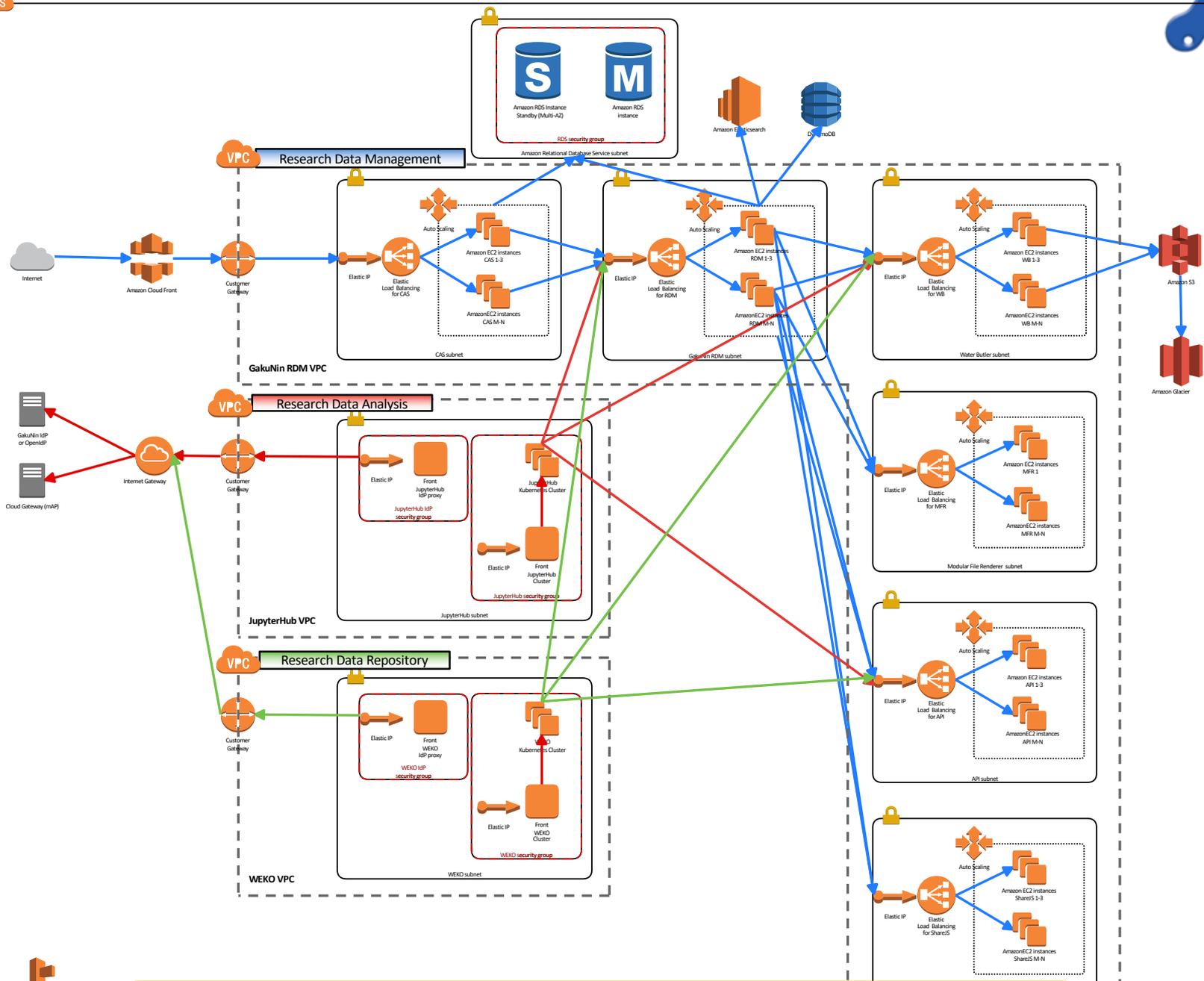
Near future (2017-2021)



System access diagram of GakuNin RDM service on SINETS and GakuNin 12

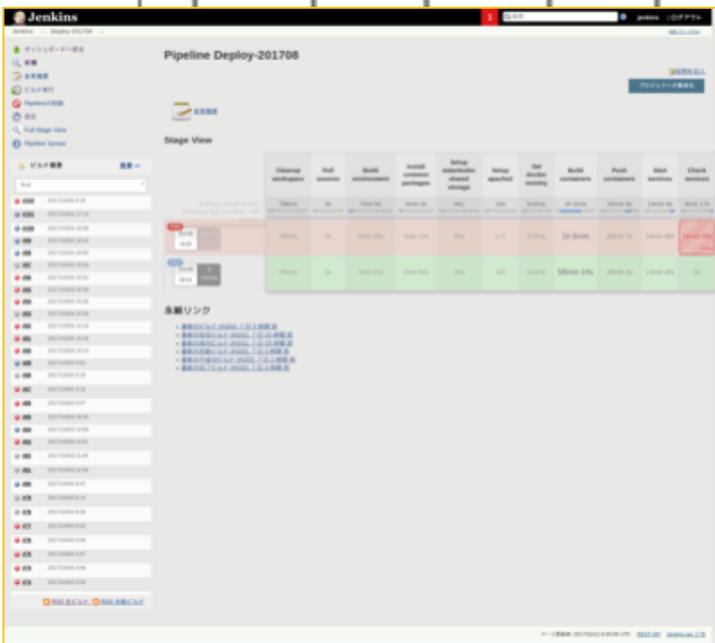
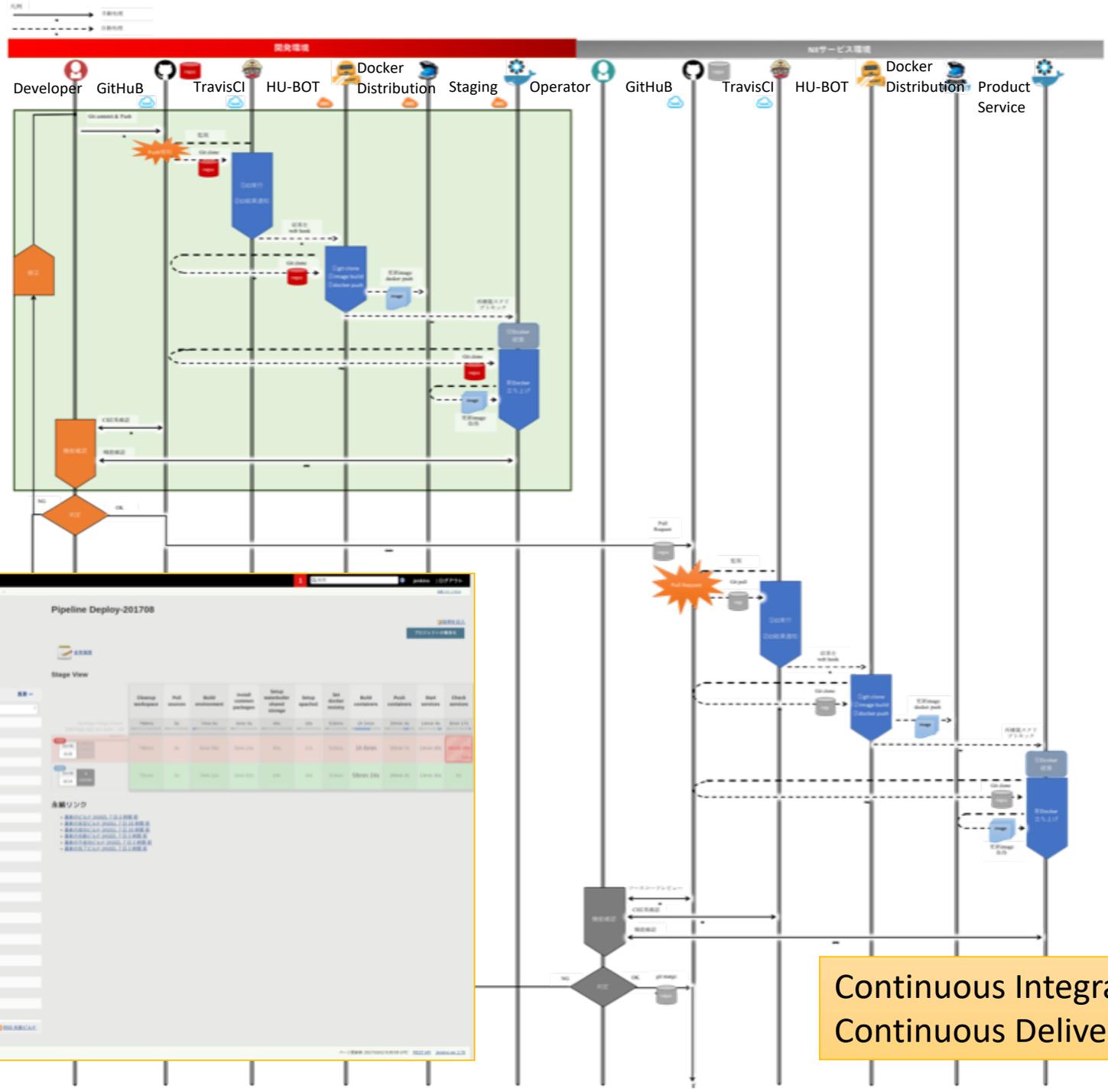
GakuNin RDM is constructed micro service architecture





Service architecture of GakuNin RDM based on micro-service





Continuous Integration and Continuous Delivery



GakuNin: nationwide authentication-federation service in Japan by NII



The screenshot shows the 'About GakuNin' page on the website. It includes the following text:

Development of the Academic Access Management Federation in Japan (GakuNin)

With the cooperation of universities etc. throughout Japan, full-scale development and management of the Academic Access Management Federation in Japan (nickname: GakuNin) started in 2009.

About the Academic Access Management Federation in Japan

The Academic Access Management Federation in Japan is a federation consisting of universities, who are users of academic e-resources, and organizations like publishers, who are providers of such e-resources. By mutually trusting rules/policy stipulated by the Federation, organizations will be able to utilize federated access between each other.

Once the federated authentication is put into place, it will enable on-campus Single Sign-On (the mechanism where a single ID and password permit a user to access all systems). It will also create an environment where a user can access other university and commercial services using a single password and without the need to re-enter the ID or password. For example, a user can use another university's wireless LAN with the home university's ID and password, and seamlessly access e-journals that the home university subscribes to.

Infrastructure of GakuNin

The diagram illustrates the infrastructure of GakuNin. At the top, Service Providers (SP) include 'E-Journal (e.g. CINA)', 'Issue certificate (e.g. Server-soft)', 'Issue an account (e.g. wireless LAN)', 'e-Learning', and 'On-campus system'. These connect to the 'GakuNin Governing Structure' which includes 'Policy', 'System Definition', and 'Rules'. Below this is the 'DS (Discovery Service)' with 'Metadata Repository' and 'Registration System'. At the bottom, Identity Providers (IdP) include 'university', 'academic society', and 'academic society'. Users like 'staff', 'student', and 'creator' are shown with their authentication process. A box notes 'Federation using Shibboleth and PID' which 'reduces the number of man hours relating to ID management' and 'enables safe and secure access management, which protects personal information'. A final note says 'Smooth access via Single Sign-On' and 'Flexible and secure access off-campus and from other universities'.

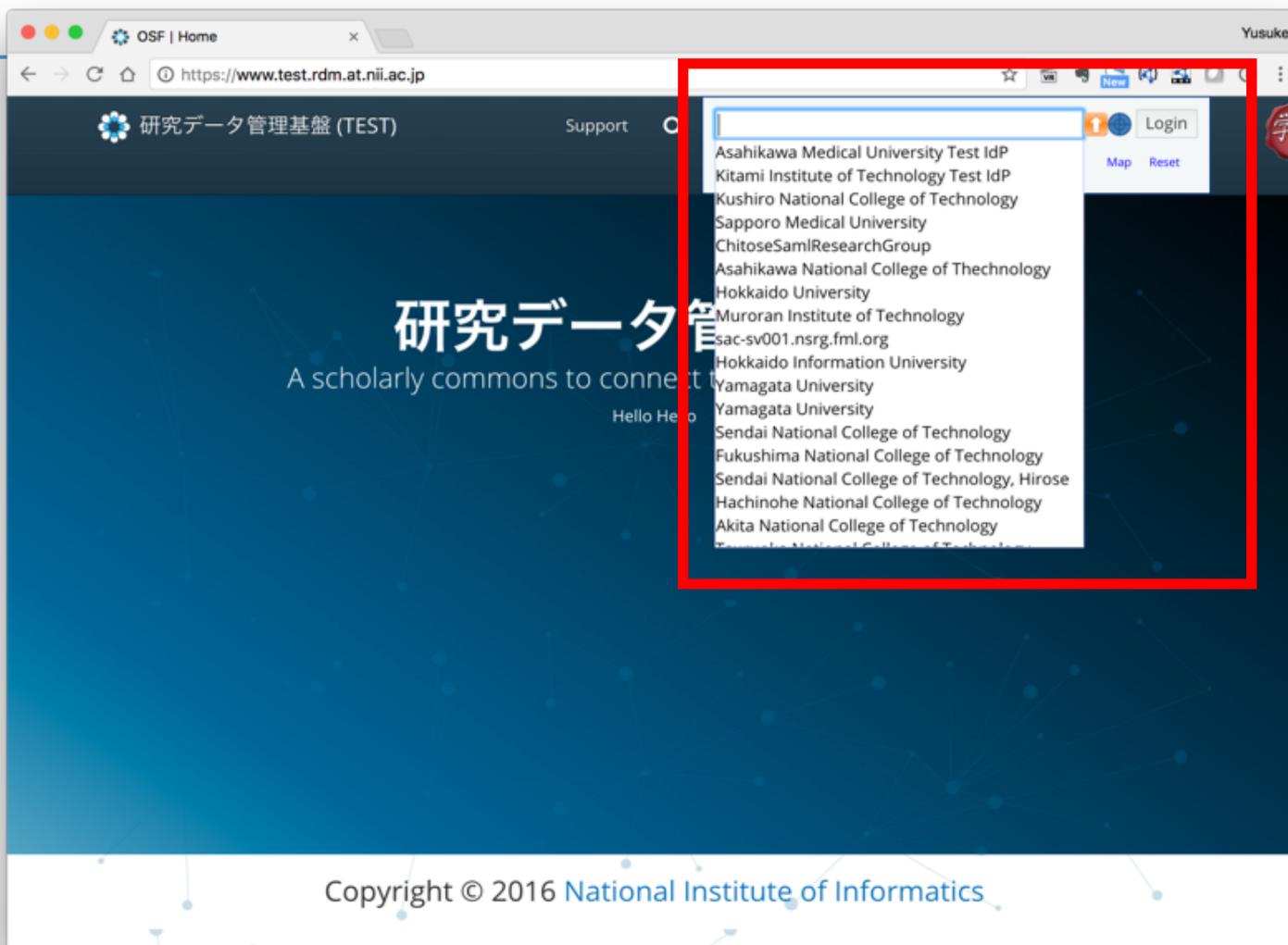
Figure: Structure of GakuNin

The technology we use is Shibboleth (Note 1), which was developed by Internet 2 (USA). Shibboleth has been used in many other countries, however, there is a wide range of technical and operational aspects that have to be decided upon, such as support for Japanese language, construction of a governing structure, and establishing associated rules. In particular, there are many aspects that have to be developed specifically for Japan, such as complying with Japanese Personal Information Protection laws.

In order to materialize these aspects, we conducted a feasibility study in 2006, implemented a pilot operation in 2009, and, based on the accumulated results, started a full-scale operation in 2010. An overview of the full-scale operation is as follows.

- It is based on Security Assertion Markup Language (SAML) and Shibboleth architecture.  Shibboleth.
- Web service developer can combine GakuNin components (IdP, SP, DS, mAP) to their service easily.
 - IdP: identify provider
 - SP: service provider
 - DS: discovery service
 - mAP: member attribute provider (for group making)
- End-user can use many SP supported service under the single sign on (SSO) securely.
- Administrator can design to trace an user attributions selectively.

GakuNin RDM (Research Data Management) service



The screenshot shows a web browser window displaying the GakuNin RDM service website. The URL is <https://www.test.rdm.at.nii.ac.jp>. The page title is "研究データ管理基盤 (TEST)". A search dropdown menu is open, listing various institutions and organizations. The menu items are:

- Asahikawa Medical University Test IdP
- Kitami Institute of Technology Test IdP
- Kushiro National College of Technology
- Sapporo Medical University
- ChitoseSamiResearchGroup
- Asahikawa National College of Thechnology
- Hokkaido University
- Muroran Institute of Technology
- sac-sv001.nsrq.fmi.org
- Hokkaido Information University
- Yamagata University
- Yamagata University
- Sendai National College of Technology
- Fukushima National College of Technology
- Sendai National College of Technology, Hirose
- Hachinohe National College of Technology
- Akita National College of Technology
- Sendai National College of Technology

The page also features a "Login" button and a "Map Reset" button. The footer contains the text "Copyright © 2016 National Institute of Informatics".



GakuNin DS

NII developers combined GakuNin (federation service of academic access-management) DS (discovery service) on GakuNin RDM.

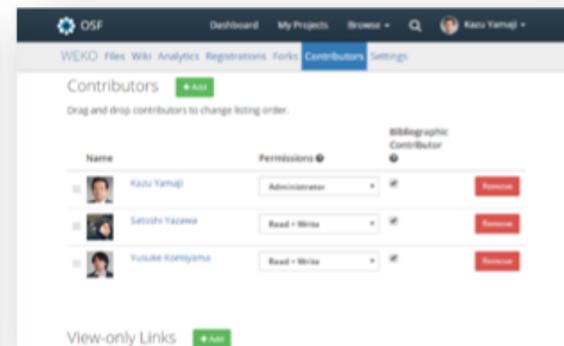
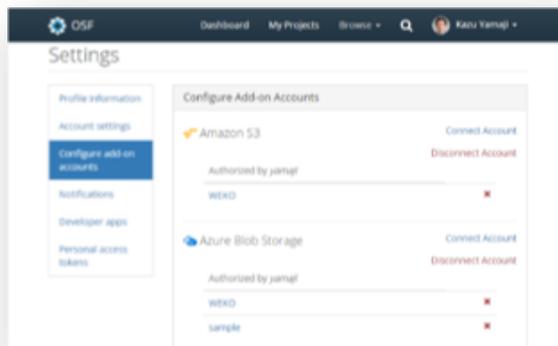
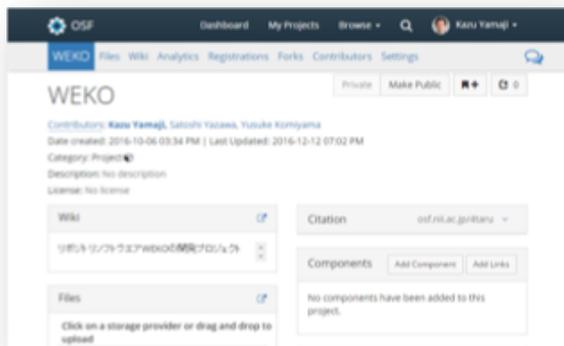
GakuNin RDM

New service!

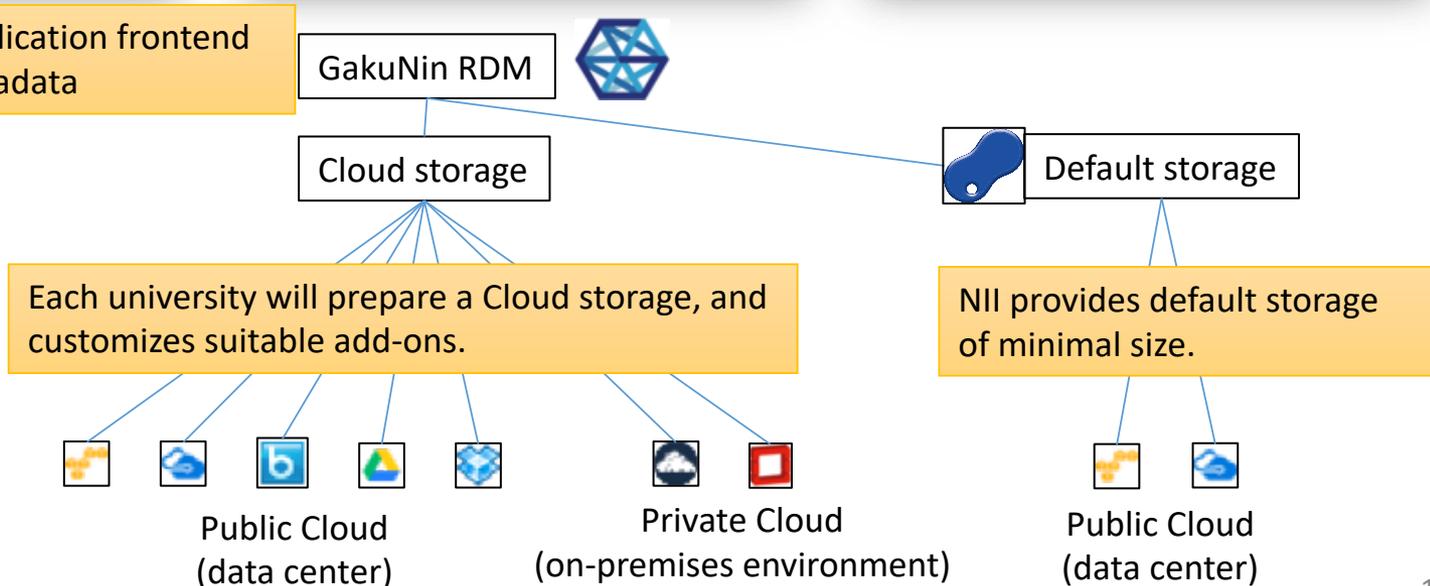
End users can manage research data on a project basis.

Cloud storage can be connected to the project workspace with various add-ons.

End users can share files with members of virtual organization with GakuNin linkage.



NII provides Web-application frontend and database for metadata



Each university will prepare a Cloud storage, and customizes suitable add-ons.

NII provides default storage of minimal size.

GakuNin Cloud

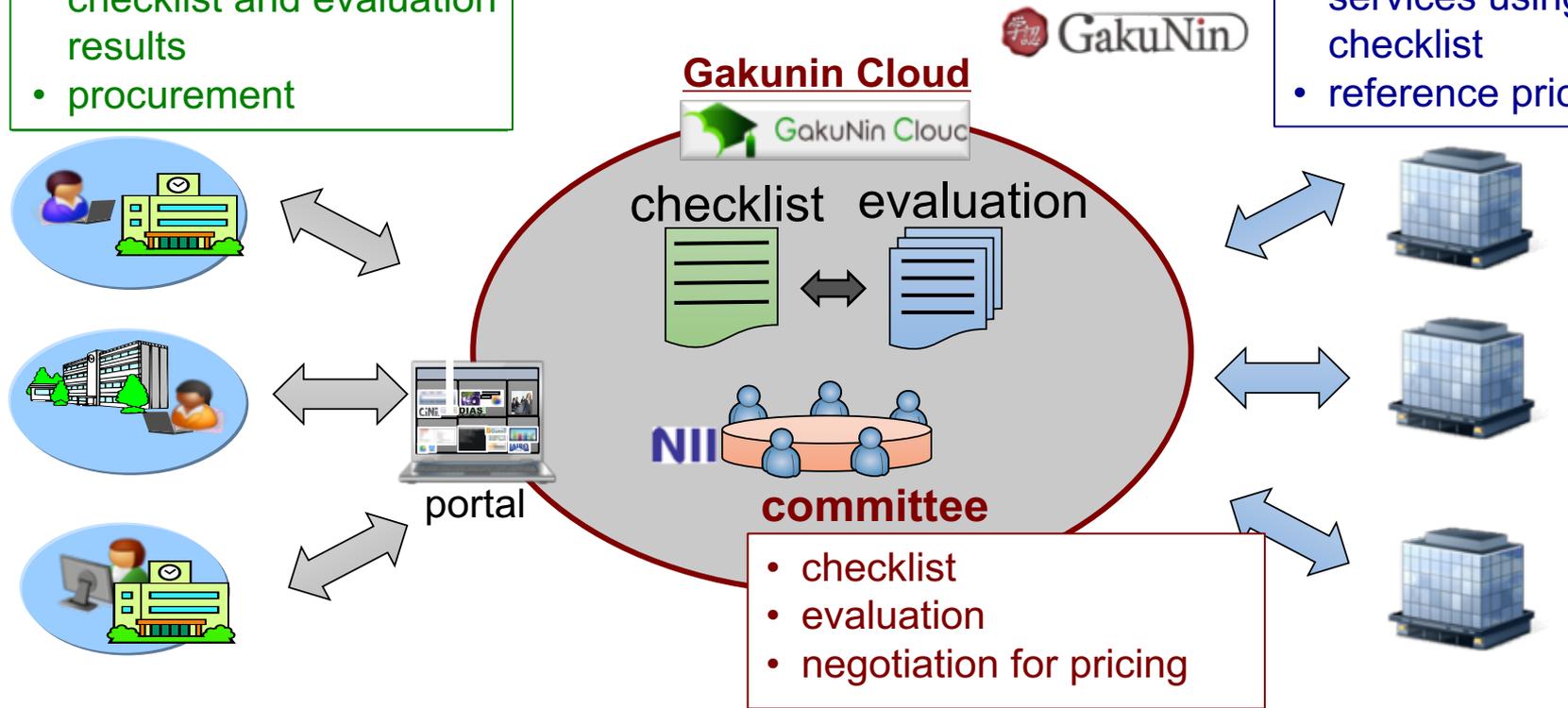
- NII helps universities/research institutes start to use cloud services:
 - checklist for cloud services and evaluation using the checklist
 - negotiation for pricing

universities/research institutes

- making spec. using the checklist and evaluation results
- procurement

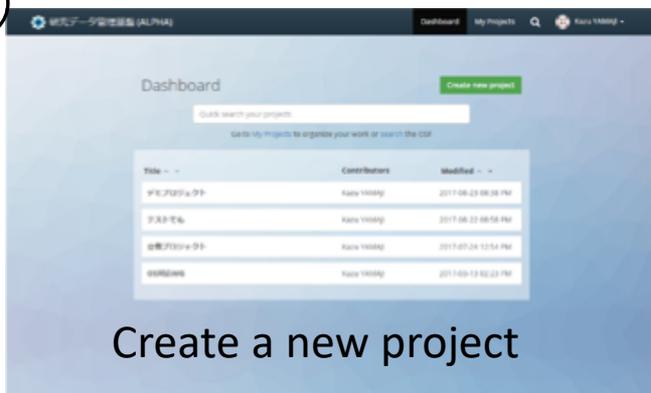
cloud providers

- evaluation of services using checklist
- reference price list



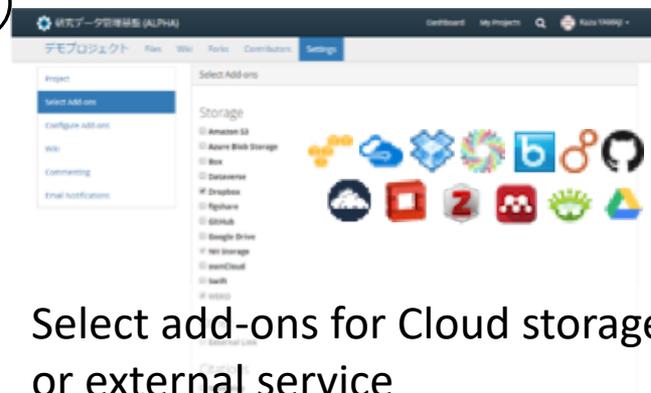
Workflow of GakuNin RDM

①



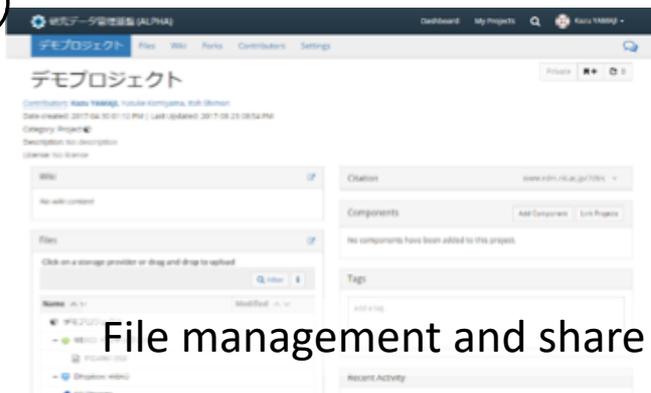
Create a new project

②



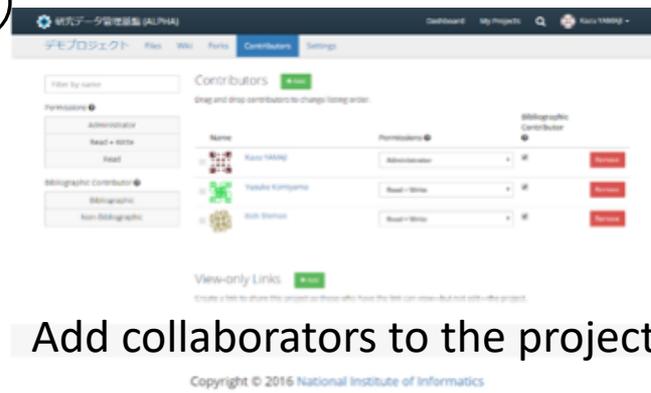
Select add-ons for Cloud storage or external service

④



File management and share

③



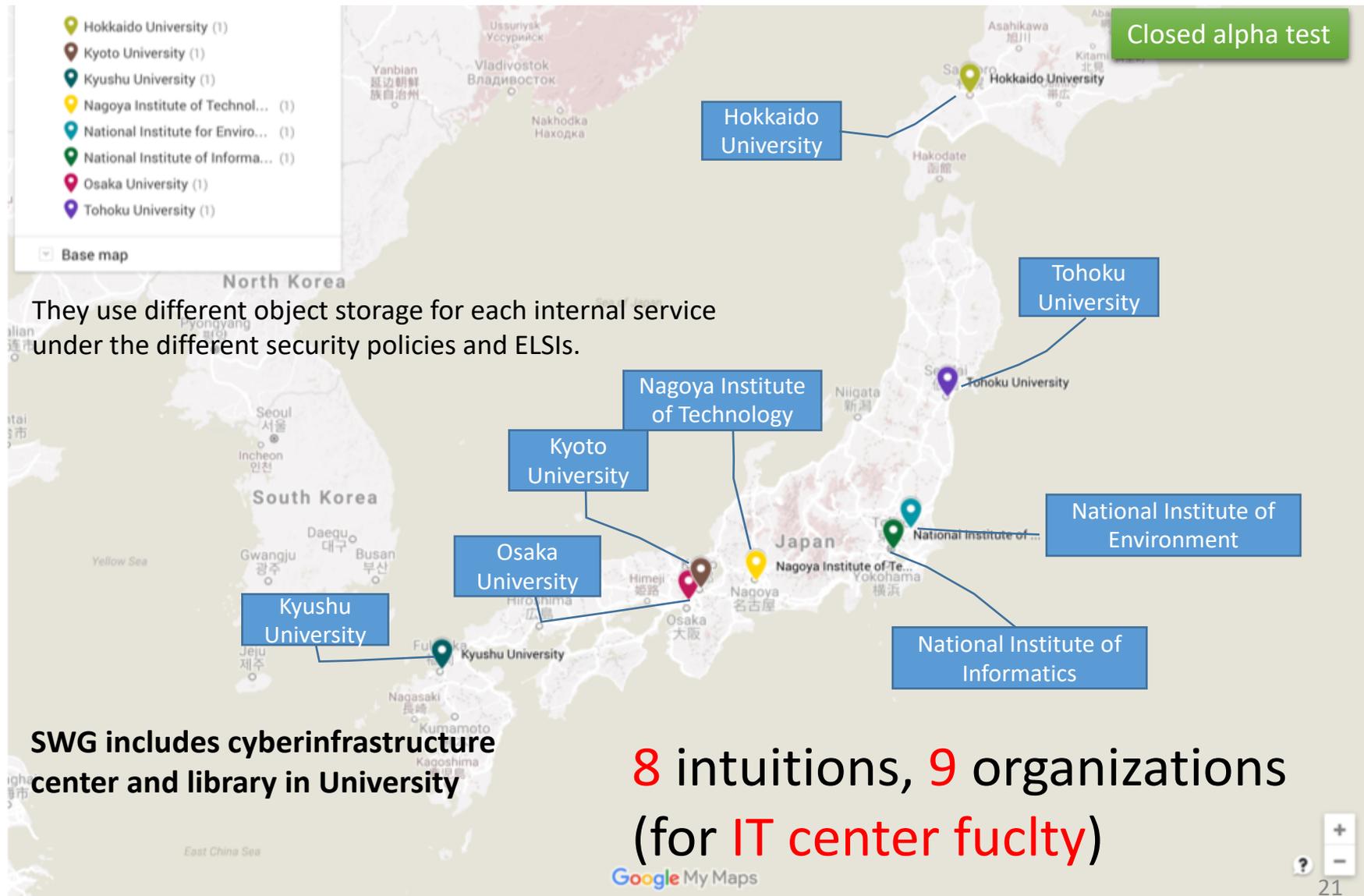
Add collaborators to the project

First closed alpha test of GakuNin RDM service

(23, Feb. — 17, Mar. 2017)

Sub Working Group (SWG) of

“working-group of advisory committee of university IT for open science”



They use different object storage for each internal service under the different security policies and ELSIs.

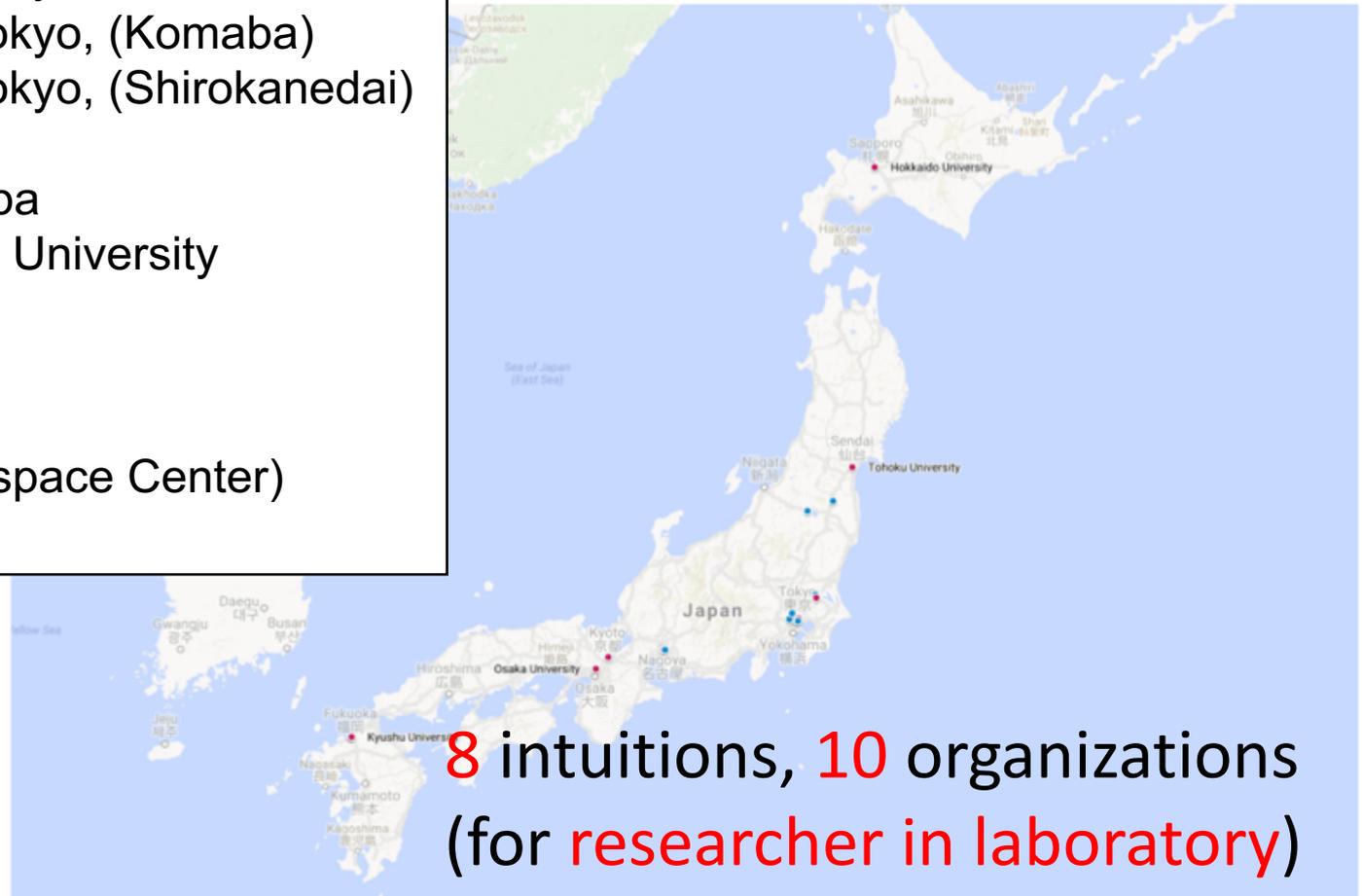
SWG includes cyberinfrastructure center and library in University

8 intuitions, 9 organizations (for IT center fuctly)

Second closed alpha test of GakuNin RDM for researcher

(12, Oct. — 12, Nov. 2017)

- The University of Tokyo,
- The University of Tokyo, (Komaba)
- The University of Tokyo, (Shirokanedai)
- Nagoya University
- University of Tsukuba
- Fukushima Medical University
- University of Aizu
- Keio University
- RIKEN
- JAXA, (Chōfu Aerospace Center)



8 intuitions, 10 organizations
(for researcher in laboratory)

Communication Wiki

OS-SWG

- ページ
- ブログ
- スペース ショートカット
- ここでチームやプロジェクトの最も重

OS-SWG ホーム

1 作成者 : confluence-admin 最終編集者 : komiyama 最終編集日 : 2017/02/22 (変更を表示)

NII研究データ管理基盤 (以下, 管理基盤) へようこそ

こちらは, **国立情報学研究所 (National Institute of Informatics: NII) 研究データ管理基盤事務局 (以下, 事務局)** です。

ここでは現在, 『大学の情報環境の在り方検討会オープンサイエンス対応WG委員』のサブワーキンググループ (以下, SWG) の皆さまへ

NII研究データ管理基盤のクローズドαテスト への参加方法をご案内しています。

また, SWGメンバーで情報を共有するためのスペースです。

GakuNin Cloud gateway (Virtual Organization service portal)

devfinch.jp, FaMOUs, Fshare, meamail, meatwiki, しばすけ, サイバーセキュリティ, シンクドメインサービス, ネットワーク管理, 研究クラウドFAC, 研究クラウドQu, 情報科学大学, 情報科学大学, 情報科学大学

Single Sign On with GakuNin Federation

NII研究データ管理基盤クローズドαテスト

開催期間
平成29年2月23(木) 17:00頃 ~ 3月17日 ()

詳細はNII研究データ管理基盤クイックス

- *期間中, 3月1日(水), 8日(水), 15日(水) 予めご了承下さい。

collect feedback comments

Pages / OS-SWG ホーム
各大学からのコメント

Added by yamaji@openidp.nii.ac.jp, last edited by yamaji@openidp.nii.ac.jp on 2017/02/01 (view change)

- 九州大学
- 京都大学
- 北海道大学
- 名古屋工業大学
- 国立環境学研究所
- 大阪大学
- 東北大学

GakuNin RDM

研究データ管理基盤
A scholarly commons to connect the entire research cycle

Copyright © 2016 National Institute of Informatics

Shibboleth-Scheduler

しばすけ Shibboleth-Schedule

shibboleth-scheduler

mailinglist

Os-swg メーリングリスト管理
全体のオプションの部

Os-swg メーリングリスト管理
全体のオプション

Remote meeting

達山悠介のパーソナル会議

遠隔会議

には

UPLOADしたデータは、学生が卒業後も、誰かがアクセス
員が他の組織に移った際のID管理が重要です。

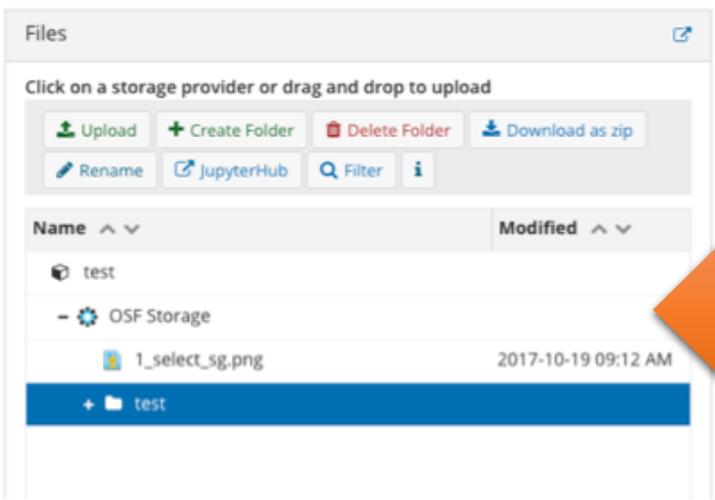
(たぶん) が、実際はあるフォルダを自動同期させる仕組み

Researcher's requirements to GakuNin RDM

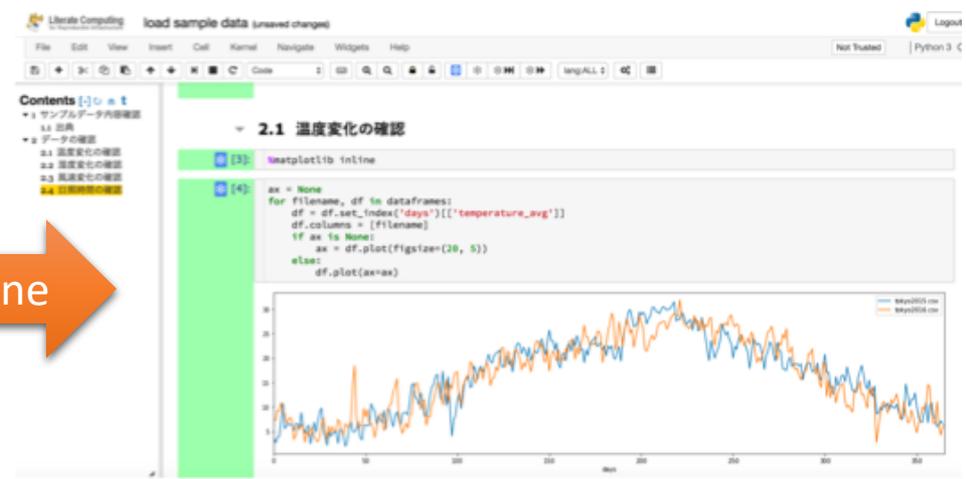
- Researchers want to..
 - combine their HPC/BigData environment to RDM.
 - make automatic data collection environment from research instrument (IoT and labo-automation).
 - write and manage a laboratory note, standard operating procedure (SOP), electronic healthcare.
 - record (EHR), data management plan (DMP) on RDM.
 - **share data analysis methods.**
 - reuse research data for education.
 - need the standard security policy and guide line about Cloud storage.
 - consider about the guide line of security policy and ethic legal social issue (ELSI).

Data analysis add-on

- We have implemented the ability to send data directly from GakuNin RDM to the statistical analysis platform to make an interactive research workflow comfortable.
- We have adopted JupyterHub as a data analysis platform and currently only Python, it can be extended to various programming languages.



RDM platform



Statistical analysis platform

Research evidence and trail management functions

Version ID	Date	User	Download	MDS	SHA2
3	2017-08-22 09:51 PM	Kazu YAMAJI	[Download]	2d66d0cc911be2f51671	f2f14e01bb309ae32c083
2	2017-08-22 09:50 PM	Kazu YAMAJI	[Download]	2fbcd150b54024971c8	8a7e7f05171e216113aa
1	2017-07-06 06:06 PM	Kazu YAMAJI	[Download]	4a86308f48a9eac04492	85095f60933a25940b6



Timestamp
By Certificate Authority

Evidence Record Syntax (ERS)

OSF Admin | Logged in as: foshim@soim.co.jp

Export institution metadata | Delete institution

国立情報学研究所[Test]

View 0 affiliated nodes

Logo: Banner:

Modify institution

- logo_name: nyu-shield.png
- banner_name: nyu-banner.png
- login_url: http://ec2-13-112-190-56-ap-northeast-1.compute.amazonaws.com/3080/Shibboleth.sso/Login?entityID=https%3A%2F%2Fshibbolethqa.es.its.nyu.edu%2Fidp%2Fshibboleth
- description: A Research Project and File Management Tool for the NYU Community: Research at NYU | Research Data Management Planning | NYU Library Research Services | Get Help
- contributors: [None]
- logout_url: http://ec2-13-112-190-56-ap-northeast-1.compute.amazonaws.com/3080/Shibboleth.sso/Logout?return=https%3A%2F%2Fshibbolethqa.es.its.nyu.edu%2Fidp%2Fprofile%2FLogout
- email_domains: [None]
- domains: [u'test-01-nyu.ccs.io']
- is_deleted: False

Project trail

Institution logs

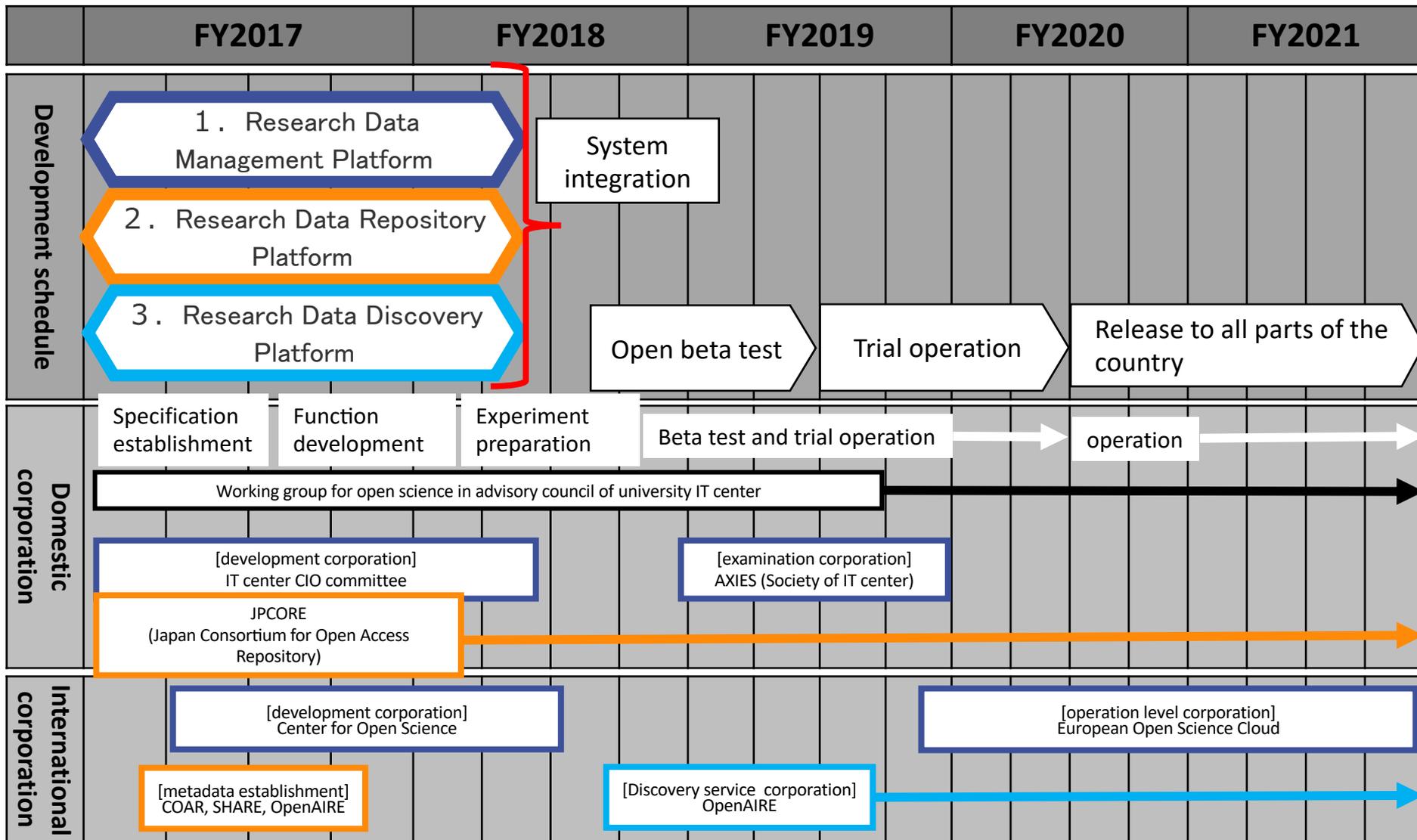
```

<?xml version="1.0" encoding="UTF-8"?>
<ns:timestamp xmlns:ns="http://www.oasis-open.org/2001/XMLSchema#dateTime" type="dateTime">
  2007-11-10T00:00:00Z
</ns:timestamp>
<ns:signature xmlns:ns="http://www.oasis-open.org/2001/XMLSchema#base64Binary" type="base64Binary">
  ...
</ns:signature>
</ns:timestamp>
  
```

XML Signature

XML Advanced Electronic Signatures (XAdES)

Roadmap of Open Science Platforms in NII



Conclusion

- NII is going to provide the nationwide research data management service, it is called GakuNin RDM. for Japanese university and research institutions for research integrity and open science.
- GakuNin RDM adopted Open Science Framework as the base system; its framework can connect external Cloud services and research tools. Cloud storage can connect to the system via SINET. This service can flexibly enhance new functions by micro service.
- GakuNin RDM can easily make corroborative group environment over the sections and institutions, and it is constructed by GakuNin federation.

Conclusion

- First closed alpha test (completion)
 - We tried to operate the pilot system of GakuNin RDM on a Cloud environment with **eight universities IT centers**, and collect their feedback comments about the system development in **March 2017**.
- Second closed alpha test (completion)
 - We try to operate GakuNin RDM for **researchers in October 2017**. New **eight universities** and one institution reviewed the service. At the second Closed alpha test, we investigated collaborative research method in various disciplines and cooperation with external systems such as research instruments or analytical software.
- Third closed alpha test (planning)
 - We are planning the third trial of GakuNin RDM for **university IT centers and research institutions** in **March 2018**.
 - If you are interested please join us.
 - rcos-office@nii.ac.jp

Thank you for your attention