Case Study in National Level Cloud Service for Institutional Repositories

Kazu Yamaji

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

Unisa Library and Information Services Open Access Webinar, 26 October 2021

2015-10-12

2015-10-18

3

2021-10-18

What is
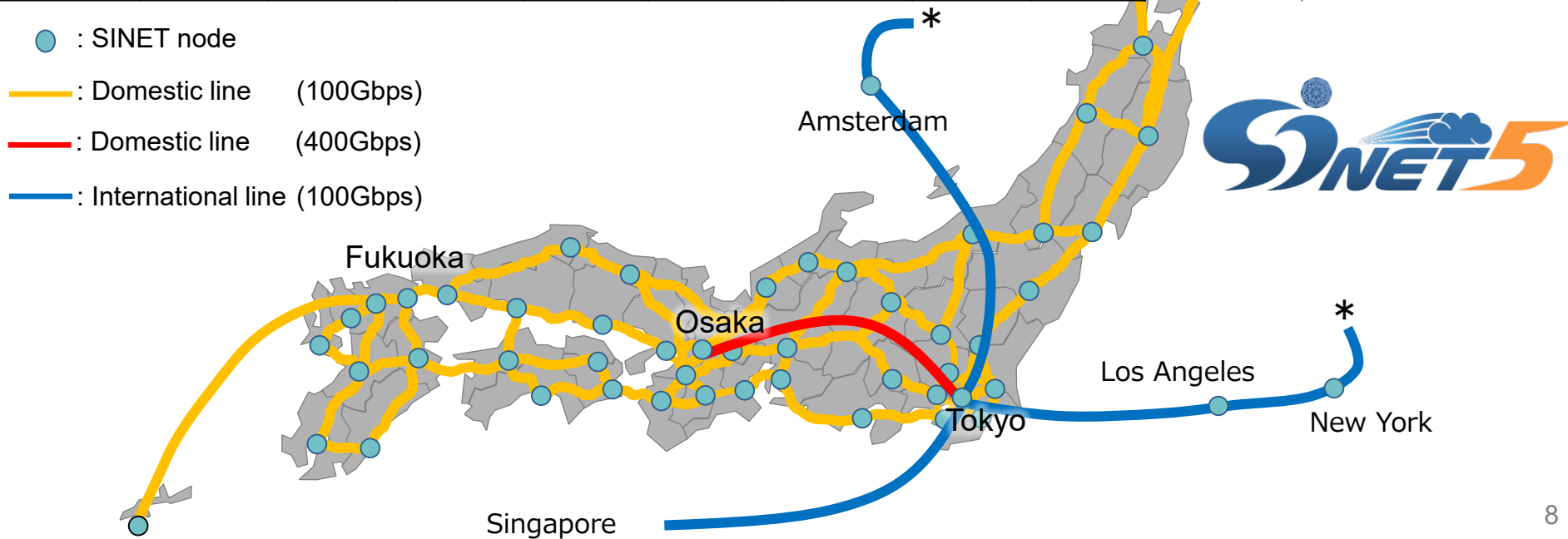# **N**ational **I**nstitute of **I**nformatics?
NII

National Research and Education Network

- SINET is a Japanese academic backbone network connecting nearly 1,000 universities and research institutions with about 3 million users.

- Current Network : SINET5 started full-scale operation in April 2016.

As of Feb. 2021

	National Universities	Municipal Universities	Private Universities	Junior Colleges	Colleges of Technology	Inter-Univ. Research Institutes	Labs and Others	**Total**
Number of Organizations	86 (100%)	86 (91%)	421 (68%)	82 (25%)	56 (98%)	16 (100%)	222	**969**

⬤ : SINET node

━━ : Domestic line (100Gbps)

━━ : Domestic line (400Gbps)

━━ : International line (100Gbps)

Sapporo

Amsterdam *

Fukuoka

Osaka

Tokyo

Los Angeles

New York *

Singapore

SINET 5

SINET5

21st Century Academic Information Infrastructure for Advancing Open Science

Collaboration and Promotion in Research and Education

Resource

◆ **Promotion of academic information circulation and open access**

◆ **Collaborative promotion of institutional repository expansion**

CiNii

JAIRO Cloud

Federation

◆ **Collaborative enhancement of authentication between universities**

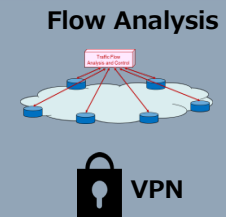GakuNin Federation

eduroam

Cloud

◆ **Dramatic cost reduction and enhancement of research and education environment by tailored cloud services**

GakuNin-Cloud Direct Connection

Security

Flow Analysis

◆ **Network flow analysis and dynamic control**
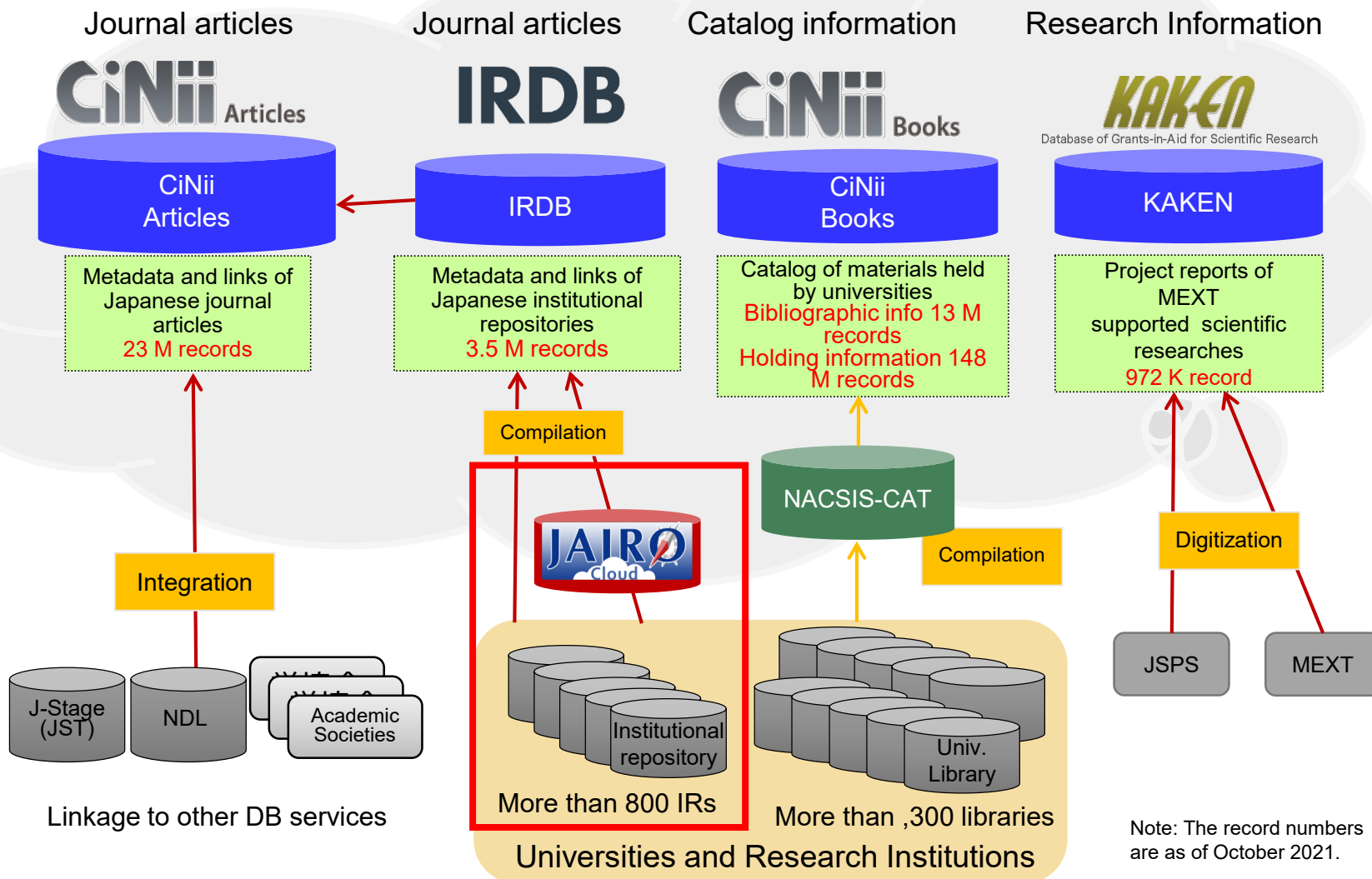
◆ **Raise of security level for SINET users**

VPN

Network

◆ **Nationwide 100-Gbps backbone network and scalable network expansion**

◆ **High-speed direct international lines to USA, Europe, and Asia**

◆ **Introduction of new technologies such as SDN in response to user needs**

Scholarly Information Infrastructure

Journal articles — **CiNii Articles**

CiNii Articles

Metadata and links of Japanese journal articles
23 M records

Journal articles — **IRDB**

IRDB

Metadata and links of Japanese institutional repositories
3.5 M records

Catalog information — **CiNii Books**

CiNii Books

Catalog of materials held by universities
Bibliographic info 13 M records
Holding information 148 M records

Research Information — **KAKEN** Database of Grants-in-Aid for Scientific Research

KAKEN

Project reports of MEXT supported scientific researches
972 K record

Compilation

Integration

JAIRO Cloud

NACSIS-CAT

Compilation

Digitization

J-Stage (JST)   NDL   Academic Societies

Institutional repository

Univ. Library

JSPS   MEXT

Linkage to other DB services

More than 800 IRs

More than ,300 libraries

Universities and Research Institutions

Note: The record numbers are as of October 2021.

10

Why
Repository **Cloud** Service?

NII-funded Programs

- NII-IRP (Institutional Repositories Program) http://www.nii.ac.jp/irp/en/
  - Phase 1 : FY2005-2007
  - Phase 2 : FY2008-2009
  - Phase 3 : FY2010-2012
  - Three categories of funding
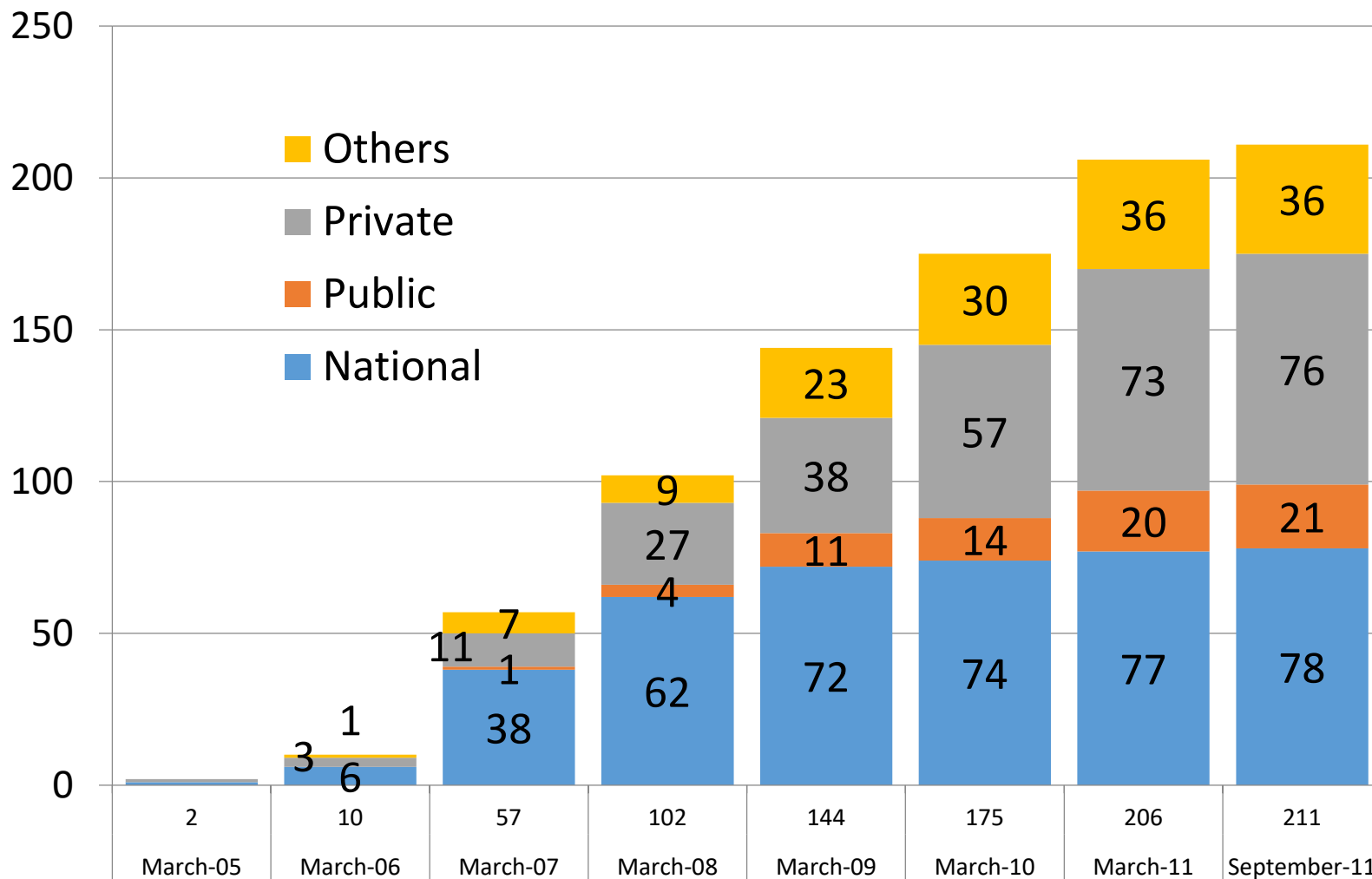   - Area 1: Support for developing IRs and content creation
    - Area 2: Research and development
    - Area 3: Support for community activities

	Phase 1			Phase 2		Phase 3		
	2005	2006	2007	2008	2009	2010	2011	2012
Area 1 (Institutions)	19	57	70	68	74	24	31	34
Area 2 (Projects)	-	22	14	21	21	8	8	7
Area 3 (Projects)	-	-	-	-	-	5	4	4

Expansion of IRs until 2011

Implementation Rate of IRs in JAPAN

**Almost all national universities have their own IRs, while the implementation rate of
other public/private universities hovers at**

20-30%

in 2012

Matters Surrounding IRs in Japan

- Ministry of education and research in Japan mandated the digitalization of doctoral dissertations and their dissemination over the network.
    - In 2013 when the above was stated, there still were about 200 universities which does not have IRs.

- There are more than 700 HEs in Japan

- The universities which have already owned their IRs were also struggling with how to update and maintain the system in the sustainable manner.
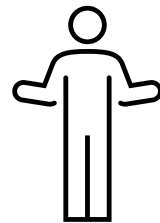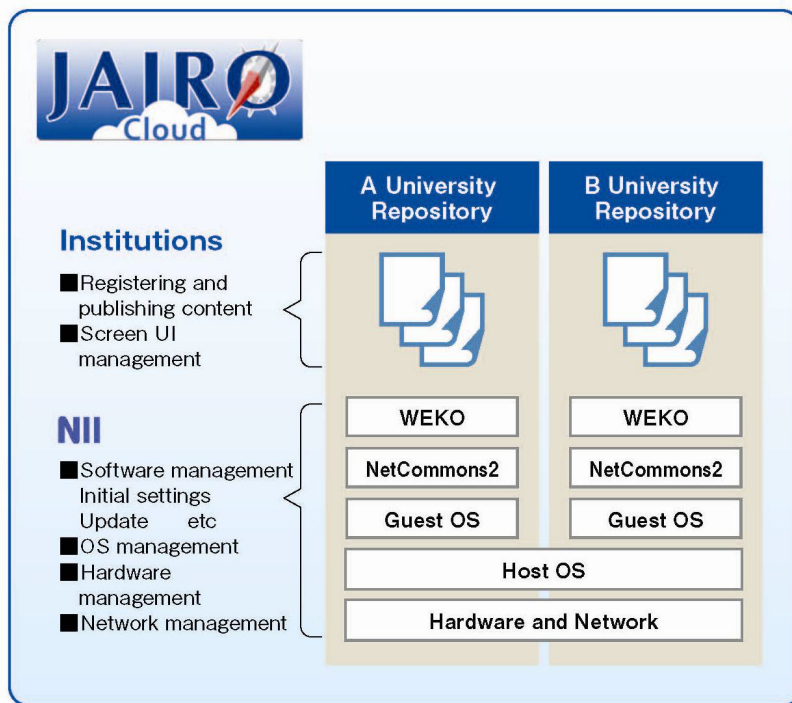
Impossible to upgrade our customized DSpace!
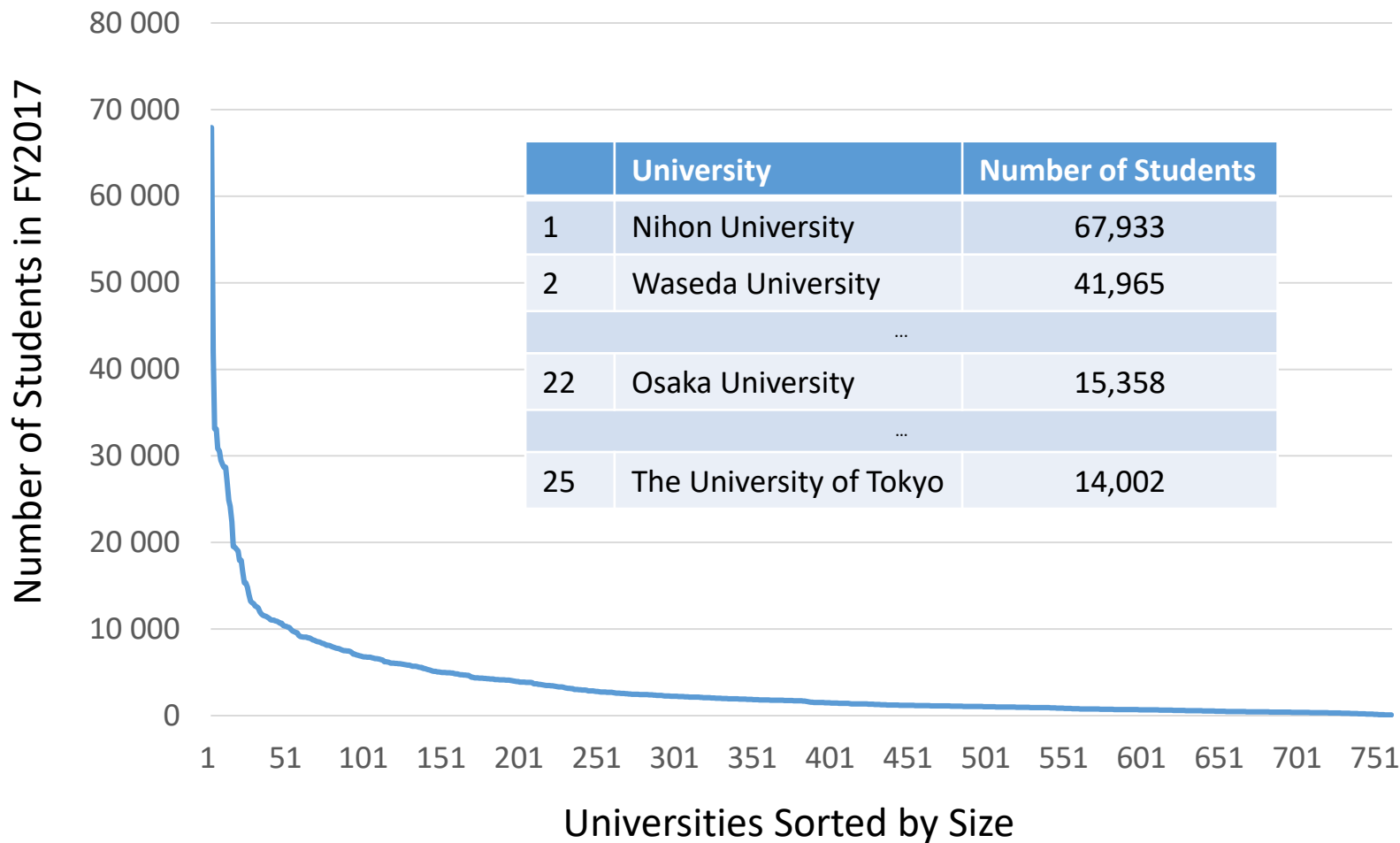
JAIRO Cloud

- Background
    - Limited resources and less technical knowledge hamper implementation of IR especially in small universities.
    - JAIRO Cloud provides a shared instance of IR system on the virtual server hosted by NII since April 2012.
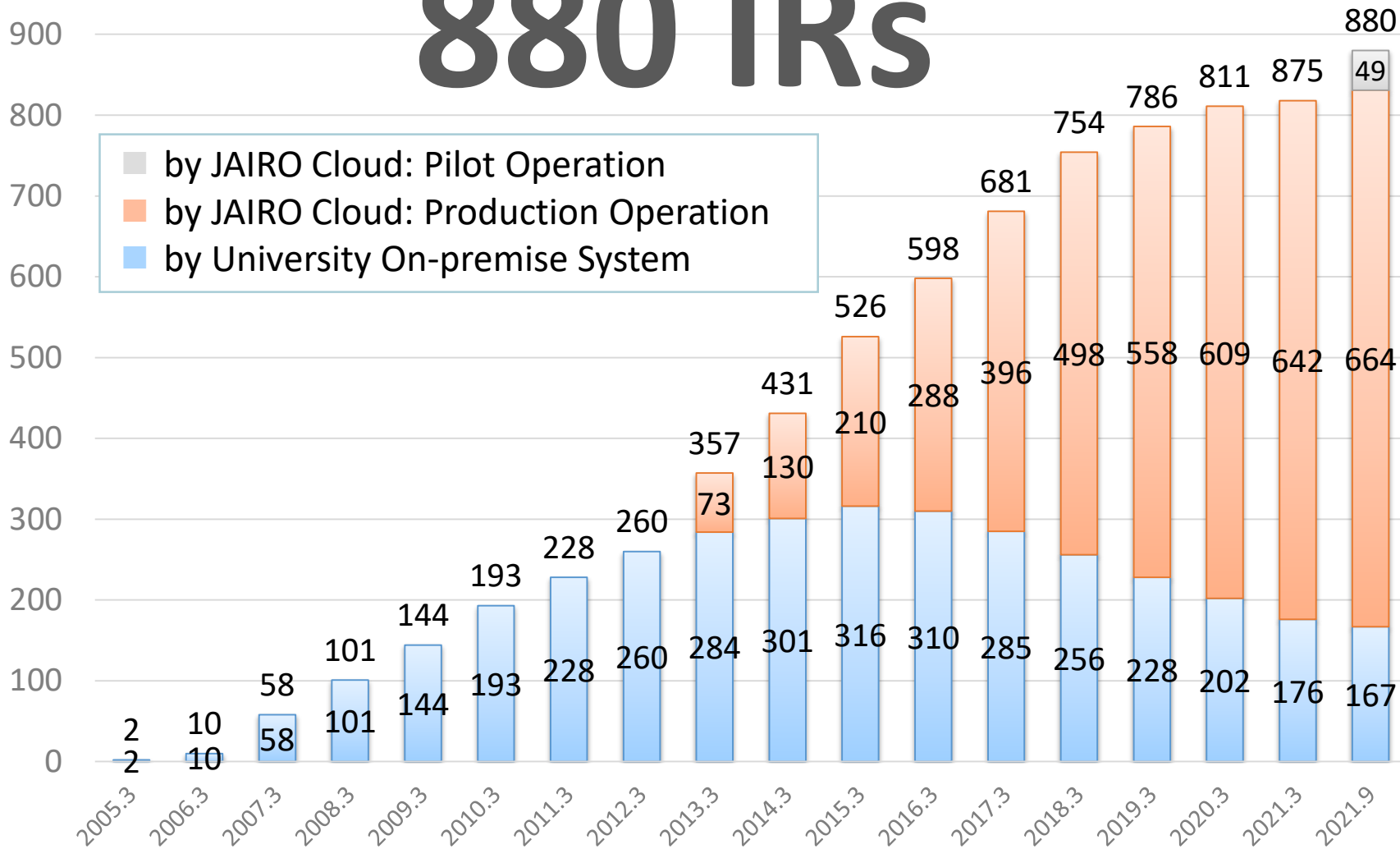
- Service Architecture

16

Long Tail Feature in Japanese Universities

	University	Number of Students
1	Nihon University	67,933
2	Waseda University	41,965
	...	
22	Osaka University	15,358
	...	
25	The University of Tokyo	14,002

Y-axis: Number of Students in FY2017

X-axis: Universities Sorted by Size

JAIRO Cloud could be a good example for supporting long-tail
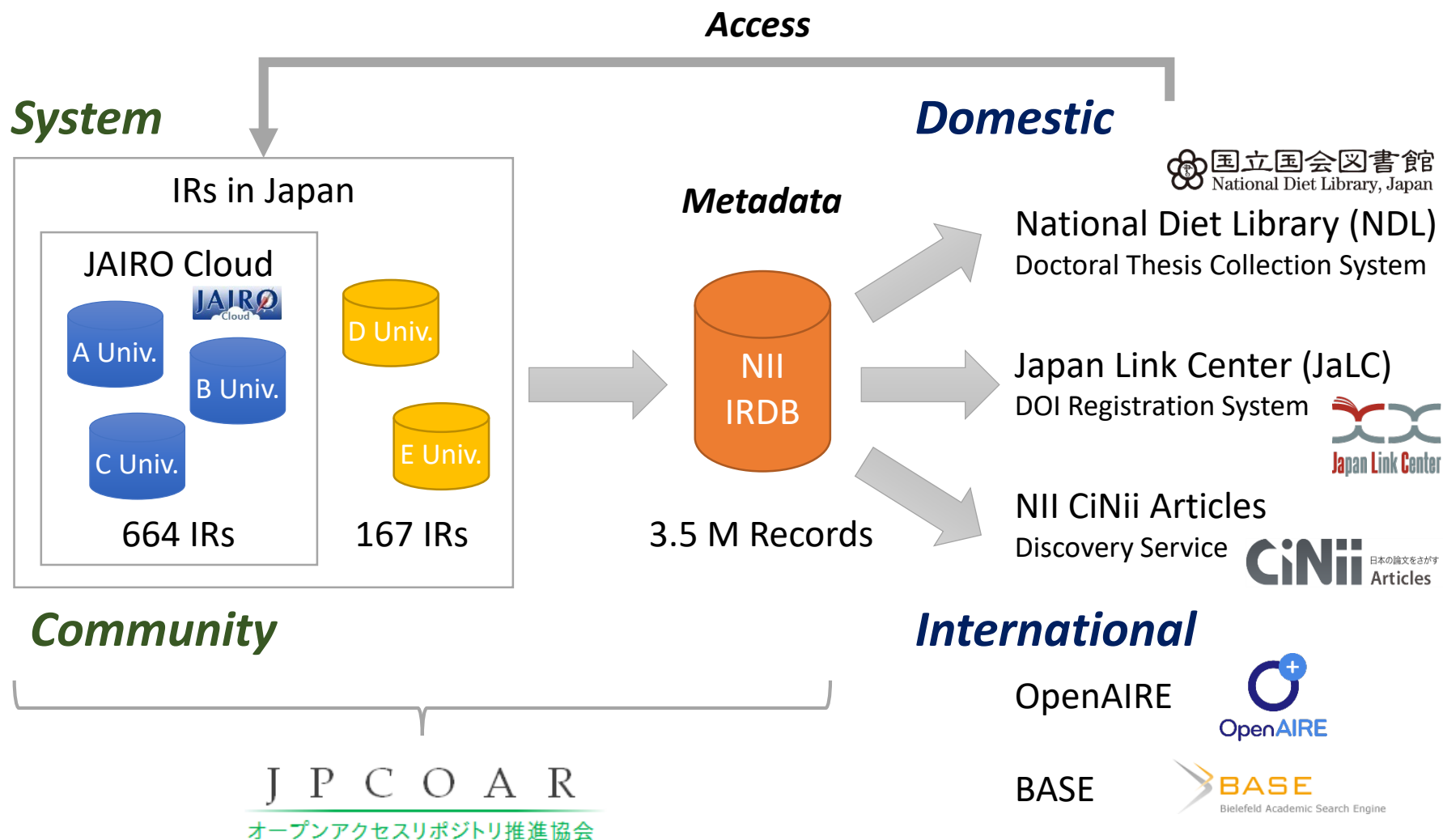
Number of Institutional Repository in Japan

880 IRs

Legend:
- by JAIRO Cloud: Pilot Operation
- by JAIRO Cloud: Production Operation
- by University On-premise System

Date	Total	Pilot	Production	On-premise
2005.3	2			2
2006.3	10			10
2007.3	58			58
2008.3	101			101
2009.3	144			144
2010.3	193			193
2011.3	228			228
2012.3	260			260
2013.3	357		73	284
2014.3	431		130	301
2015.3	526		210	316
2016.3	598		288	310
2017.3	681		396	285
2018.3	754		498	256
2019.3	786		558	228
2020.3	811		609	202
2021.3	875		642	176
2021.9	880	49	664	167

JAIRO Cloud Fee

Number of organization researchers	JPCOAR membership fee (JPY)*	JAIRO Cloud service fee(JPY including TAX)	Total cost (JPY including TAX)
0 – 100	20,000	43,200	63,200
101-200	20,000	86,400	106,400
201-300	20,000	129,600	149,600
301-400	20,000	172,800	192,800
401-500	20,000	216,000	236,00
501-600	20,000	259,200	279,200
601-700	40,000	302,400	342,400
701-800	40,000	345,600	385,600
801-900	40,000	388,800	428,800
901-1000	40,000	432,000	472,000
1001-1100	40,000	475,200	515,200
1101-1200	60,000	518,400	578,400
1201-1300	60,000	561,600	621,600
1301-1400	60,000	604,800	664,800
1401-1500	60,000	648,000	708,000
1501-	60,000	691,200	751,200

*If an institution requires to use JAIRO Cloud service, the institution should join JPCOAR with paying membership fee.

Japanese IR Network

JPCOAR

- Digital Repository Federation (DRF), since **2006**
- JAIRO Cloud Community, since **2012**
- Institutional Repository Promotion Committee, since **2013**
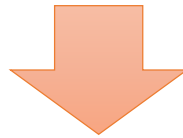
- Japan Consortium for Open Access Repository (JPCOAR), since July **2016**

JPC5OAR
2021.7.27   th Anniversary

JPCOAR: WG and TF

- **Working Group**
  - Research Data WG
  - Contents WG
  - Community Empowerment WG
  - Training WG
- **Task Force**
  - Next Mid-term Strategy TF
  - Next JAIRO Cloud TF
  - COAR Community Framework TF

JAIRO Cloud

Cloud type (SaaS) service for Institutional Repository operated by JPCOAR and NII.

JPCOAR Member

Institutional Type	Institutions	JAIRO Cloud Users
National University	80	56
Public University	70	63
Private University	416	386
Inter-University Research Institute	12	10
Junior college	49	107
National Institute of Technology College	6	6
Others (Research Institute)	45	36
Total	**678**	**664**

Research Data WG

- RDM Training Course Development

    Course 1 : Research Data Management for Open Science

    Course 2 : Design and Practice of RDM Service

    Course 3 : Research Data Management for Researcher

Provide each course from LMS
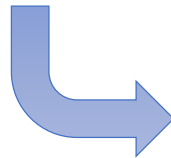
COAR Community Framework Task Force

- Objective
  - Evaluate and analyze institutional repositories in Japan by using COAR Community Framework

- Methods

Original COAR Framework
- 6.1 In cases where the repository is collecting sensitive research data, there are mechanisms that allow data owners to limit access to authorized users only

Our checklist
- 6.1.1 Does the repository cover the collection of sensitive data? ☐ Yes ☐ No
- 6.1.2 Does the repository have the ability to restrict access to resources? ☐ Yes ☐ No
- 6.1.3 If the repository has the ability to restrict access to resources, what methods can be used to control it?
  ☐ By authenticated user
  ☐ Within IP address range
  ☐ Unauthenticated users (e.g. issuing a temporary URL for peer review)
  ☐ Other

- Introduce the checklist in Asia OA webinar on April 27, 2021
  - https://www.coar-repositories.org/asia-oa/
- Ask JPCOAR working group members to conduct validation for several institutions.

Result of Evaluation

Example from Hitotsubashi University
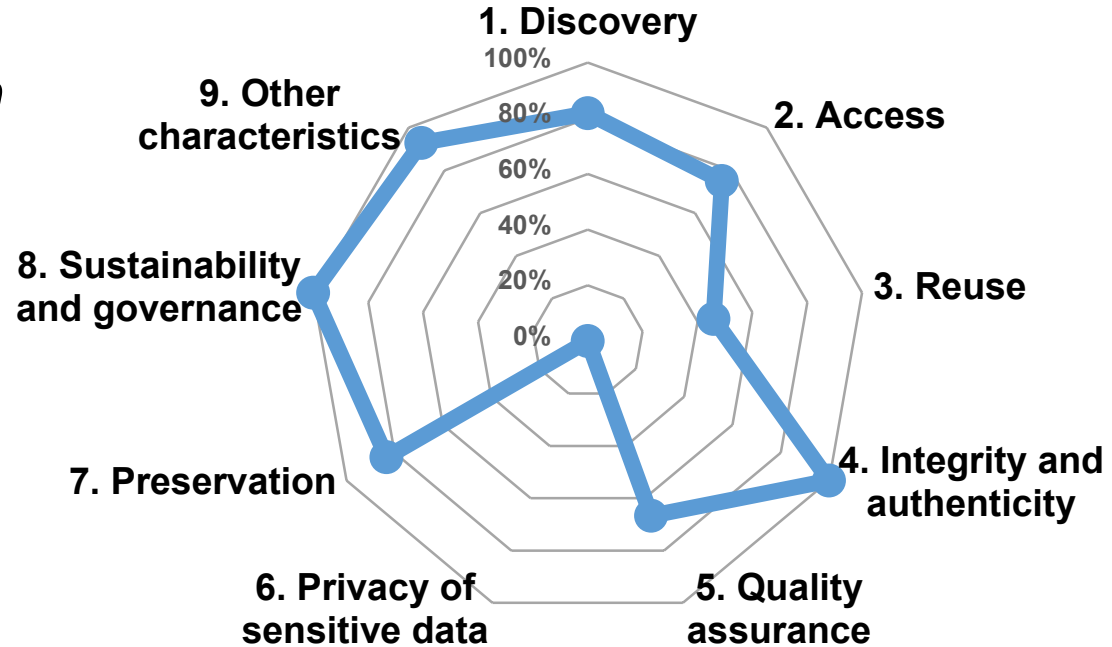
Areas for improvement

3. Our repository does not support **the display of CC licenses** and **a list of recommended file formats**, and **text and data mining**.
5. The text and metadata published in our repository do not allow for **external annotations, comments, and reviews, etc**.
6. This answer was given because the research results handled in our repository are essentially open access and we **do not deal with sensitive data in the repository**.

2005

NII Institutional Repository Project

IRP 学術機関リポジトリ構築連携支援事業

DRF Digital Repository Federation

DRF Repository Community

2013

JAIRO Cloud

2016

J P C O A R
オープンアクセスリポジトリ推進協会

NII — Inter-University Research Institute Corporation / Research Organization of Information and Systems
National Institute of Informatics

2021

Once we had a core service, strategic action can be made.

Task Force: formed to operate a time-limited project
**COAR Community Framework Task Force**

• Objective
  - Evaluate and analyze institutional repositories in Japan by using COAR Community Framework
• Challenge
  - It is necessary to arrange the framework in a format that is easy for Japanese institutions to use.
• Approach
  - Translate the Framework into Japanese
    http://doi.org/10.34477/00000534
  - Transform abstract expressions into concrete questions that can be answered in Yes/No. And create JPCOAR Checklist from the Framework

2021/10/26   Asia OA Meeting 2021   1

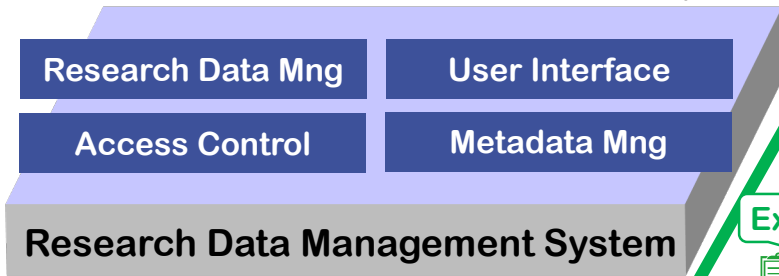研究者のための研究データマネジメント
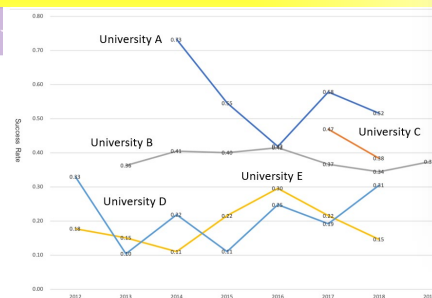
Furthe**rm**ore

Research Data Infrastructure for Open Science

CiNii Research

Discovery Platform

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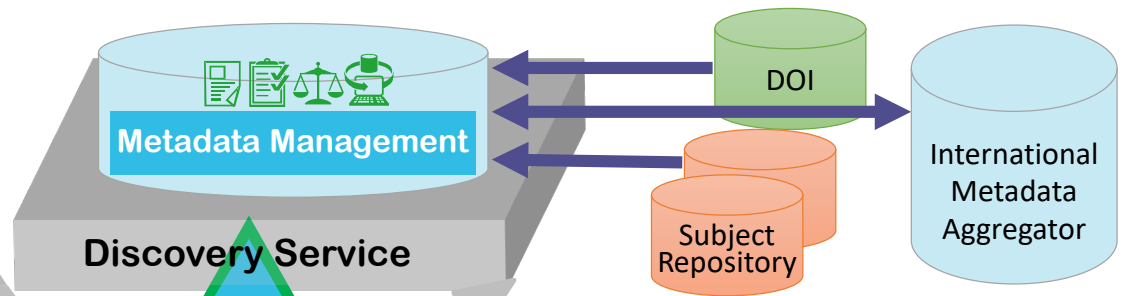- Data Exchange with International Discovery Service

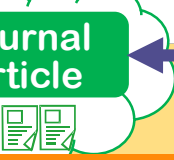**Metadata Management**

DOI

International Metadata Aggregator

Subject Repository

Discovery Service

Re-use

Metadata Aggregation

User Flow
Data Flow

Search/Find

Data User

Journal Article

Supplemental Data

Institutional Research Data Mng

Research Data Mng	User Interface
Access Control	Metadata Mng

Research Data Management System

Data Depositor

Exp/Store

Archive

Research Data Repository

RDM Platform

GakuNin RDM

Exp Data

Article

Publication Platform

JAIRO Cloud by WEKO3

Private | Shared | Public

- High Speed Access using SINET5
- Data Sharing Func using Virtual NW and ID Federation
- Effective Data Storage Switcher

Hot Storage | Hot Storage | Hot Storage

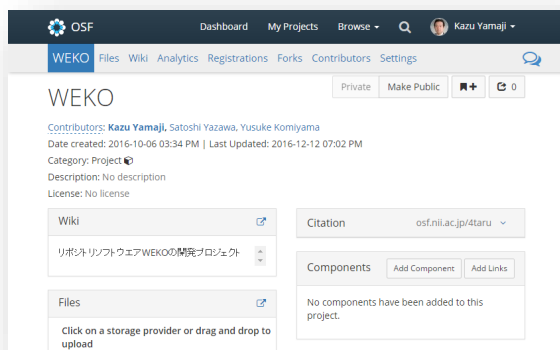Cold Storage | Cold Storage | Cold Storage
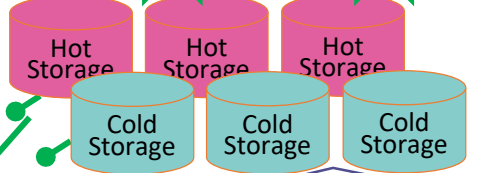
- Data oriented Self-Archiving Func
- Versioning and auto-Packaging Func
- User Dependent Personal Data Pseudonym Func

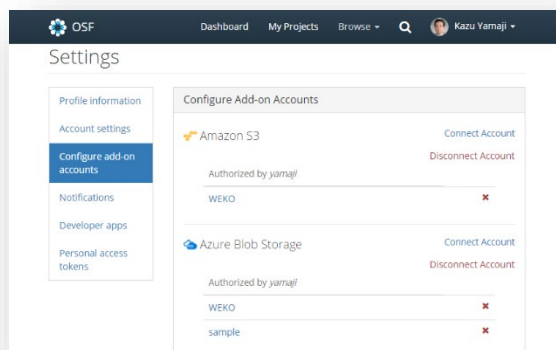Storage Area for Long-term Preservation

28

New Service GakuNin RDM

Extension of Open Science Framework developed by COS, USA

Manage Research Data by Research Project

Connect Cloud Storage from Various Plugin

Share Research Data within Collaborators Authn by ID Fed

NII: Frontend Service → RDM Platform

University: Backend Storage → Cloud Storage — NII Storage

Customize selectable Plugin depending on University Environment and Policy

Default (minimum?) Storage by NII

Public Cloud (Provider DC)

Private Cloud (On-premise)

Public Cloud (Provider DC)
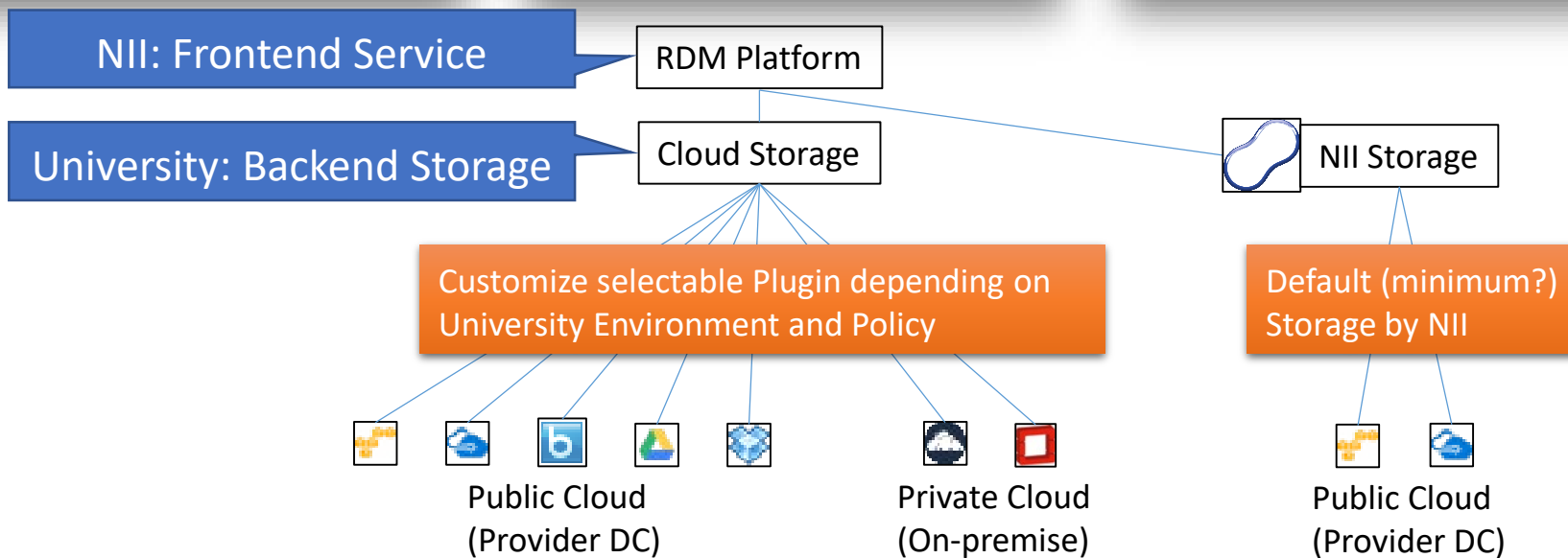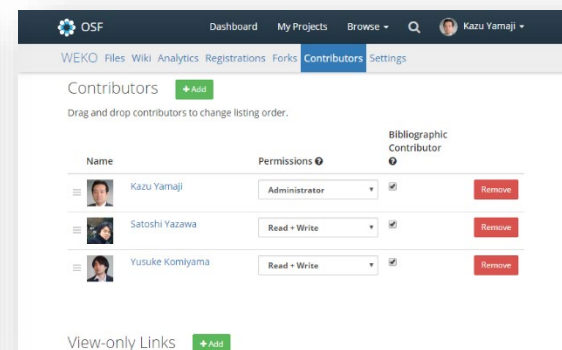
GakuNin RDM User Institutions

Connect GakuNin RDM to Repository

(for researcher and librarian)

Researcher

- File Management
- Timestamp Proofing
- Long-term Preservation
- …

Librarian, Research Office

- Metadata Management
- Data Publication
- DOI Registration
- …

Additional Services

Develop the Best Practice in the Specific Domain e.g., Social Science and Humanities
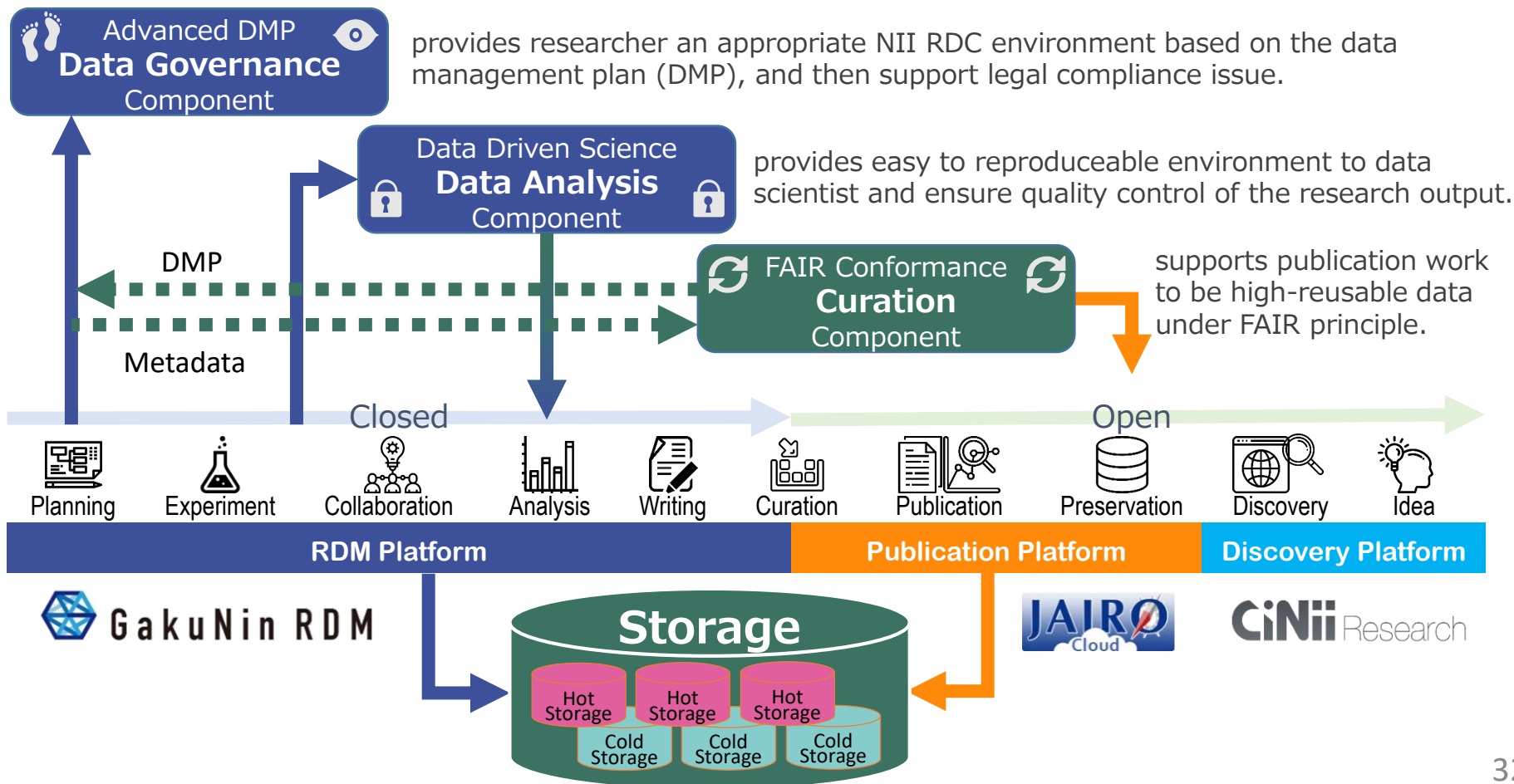
Training Platform for building Data Literacy Capacity

Educational Platform which allow users to obtain advanced RDM literacy based on NII RDC

Advanced DMP Data Governance Component — provides researcher an appropriate NII RDC environment based on the data management plan (DMP), and then support legal compliance issue.

Data Driven Science Data Analysis Component — provides easy to reproduceable environment to data scientist and ensure quality control of the research output.

FAIR Conformance Curation Component — supports publication work to be high-reusable data under FAIR principle.

DMP

Metadata

Closed — Open

Planning · Experiment · Collaboration · Analysis · Writing · Curation · Publication · Preservation · Discovery · Idea

RDM Platform | Publication Platform | Discovery Platform

GakuNin RDM JAIRO Cloud CiNii Research

Storage

Hot Storage · Hot Storage · Hot Storage
Cold Storage · Cold Storage · Cold Storage

NII RCOS — Research Center for Open Science and Data Platform

RCOS

yamaji@nii.ac.jp