

Advanced Research Data Management Services for Academia provided by the National Institute of Informatics (NII) in Japan

Yusuke Komiyama
National Institute of Informatics (JAPAN)

6, Nov. 2024

TOPIC

- **NII Research Data Cloud Vision and Current Status**
 - Background of NII Research Data Cloud
 - Work through of Research Data Platform
- **Advanced Applications**
 - Introduction of Code Package (Data Analysis) Function
 - Introduction of Data Governance (Data Management Plan) Function

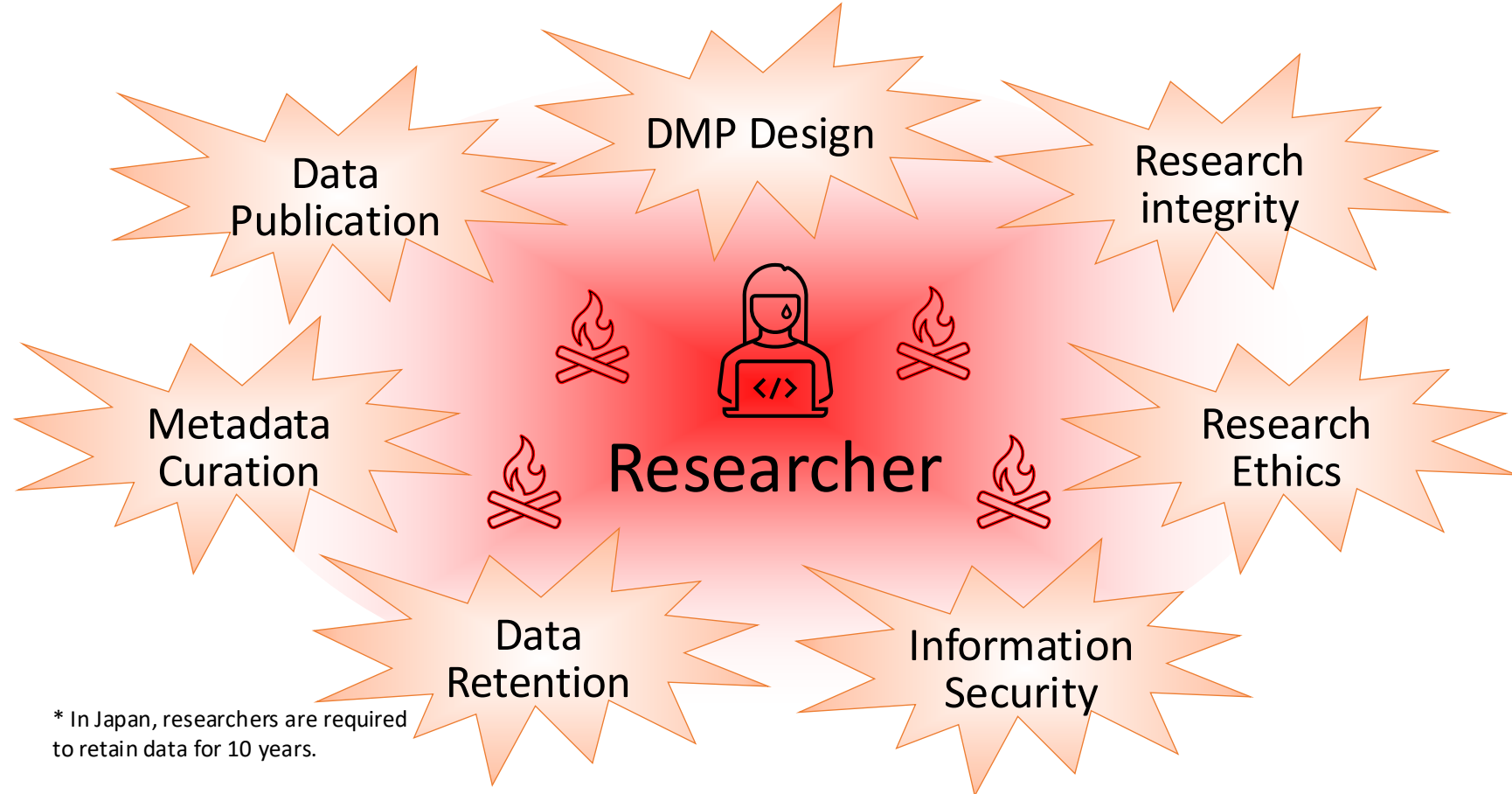
NII Research Data Cloud Vision and Current Status

Background of NII Research Data Cloud

Policy Developments on Open Science in Japan

March 2015	Cabinet Office, "Promoting Open Science in Japan" ⇒ Follow-up discussions have been take place since 2015
January 2016	"The 5th Science and Technology Basic Plan" (In Japanese) ⇒ Open Science has been promoted in order to reinforce the intellectual infrastructure
July 2016	Science Council of Japan (SCJ), "Recommendations Concerning an Approach to Open Science that Will Contributes to Open Innovation"
June 2017	Cabinet Office, "Comprehensive Strategy for Scientific and Technological Innovation 2017" (In Japanese)
June 2018	Cabinet Office, "Integrated Innovation Strategy" (In Japanese)
June 2018	Cabinet Office, "Guideline for Establishing Data Policy at National Research and Development Agencies" (In Japanese)
March 2019	Cabinet Office, "Guidelines for the Development and Operation of Research Data Repositories" (In Japanese)
June 2019	Cabinet Office, "Integrated Innovation Strategy" (In Japanese)
December 2019	Cabinet Office, "Report on the Strategy for Research Data Infrastructure Development and International Expansion" (In Japanese)
November 2019	Science Council of Japan (SCJ), "Proposal for the Development of a Sustainable Data Infrastructure for Life Sciences" (In Japanese)
May 2020	Science Council of Japan (SCJ), "Proposal for Toward Deepening and Promoting Open Science" (In Japanese)
May 2021	Cabinet Office, "6th Basic Plan for Science, Technology and Innovation" (In Japanese)
April 2021	Council for Integrated Innovation Strategy, "Fundamentals of the Management and Use of Research Data using Public Funds" (In Japanese)
December 2022	Science Council of Japan (SCJ), "Responses to deliberations on the promotion of research DX - especially from the perspective of promoting open science and data utilization -" (In Japanese)

Increasing workload for researchers towards research digital transformation

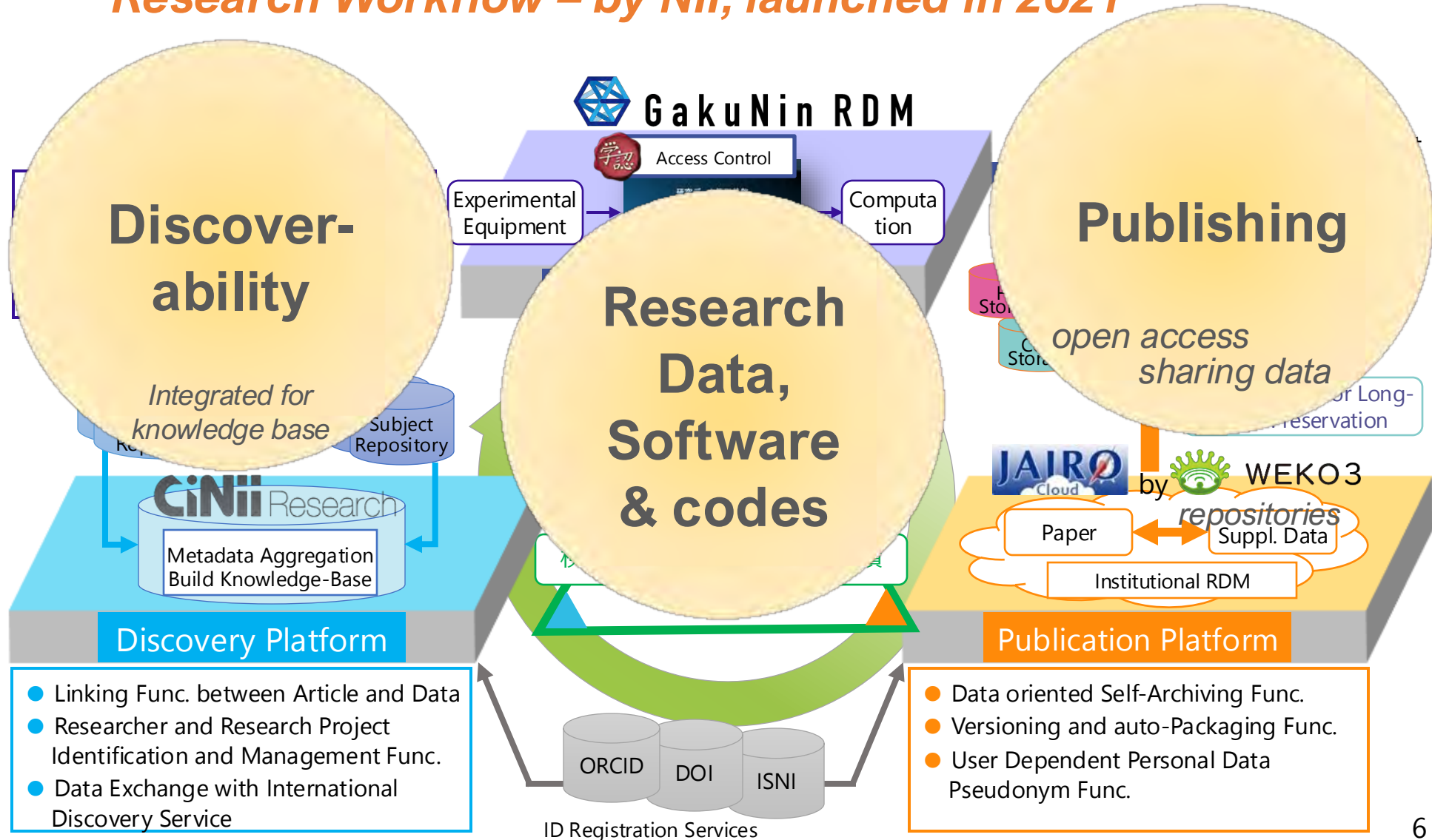


* In Japan, researchers are required to retain data for 10 years.

So much extra work to do

Positioning of GRDM as a Service Component within NII RDC ~NII Research Data Cloud – RCOS, since 2017~

Research Workflow – by NII, launched in 2021

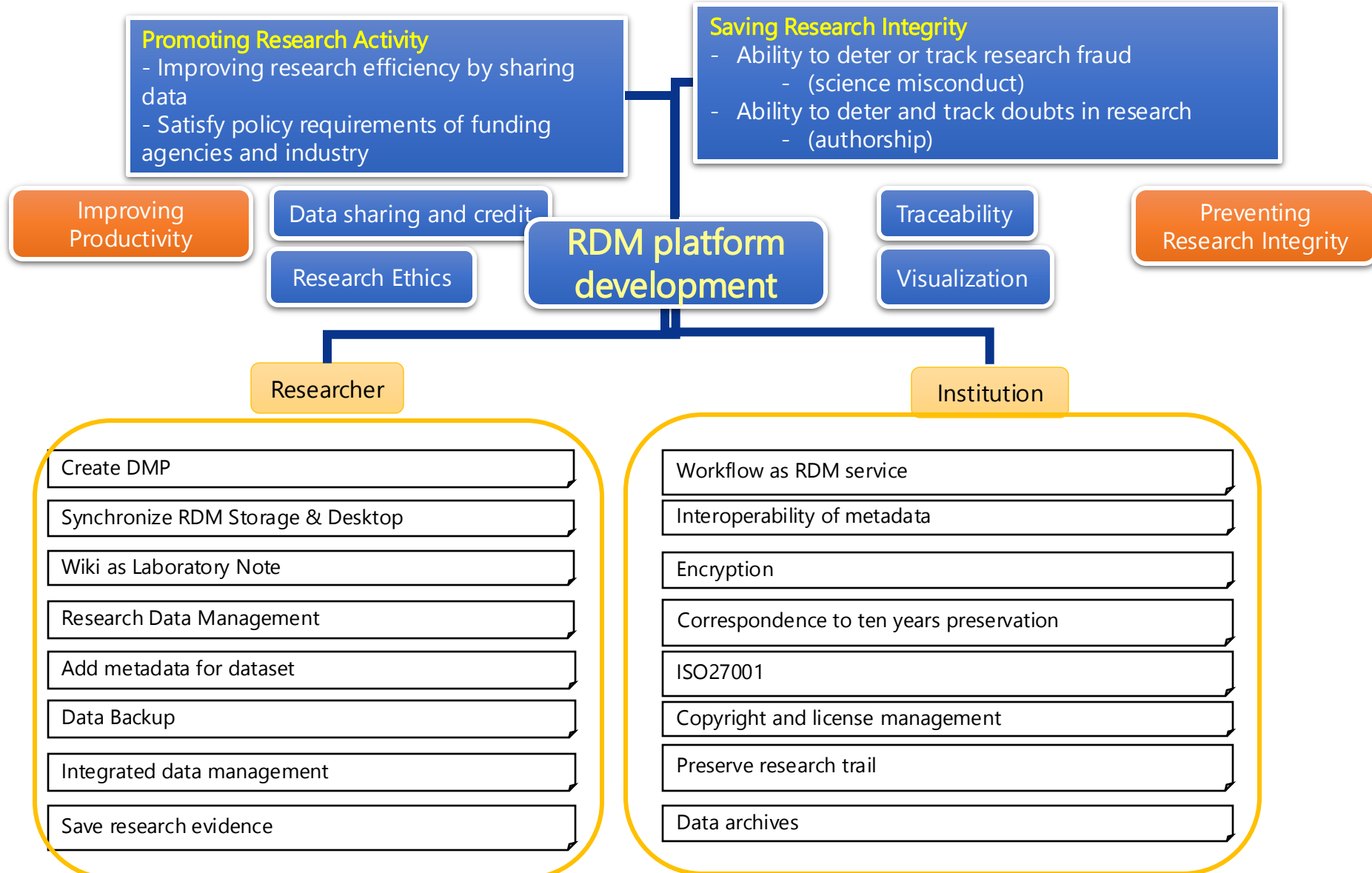


GakuNin RDM is a RDM service

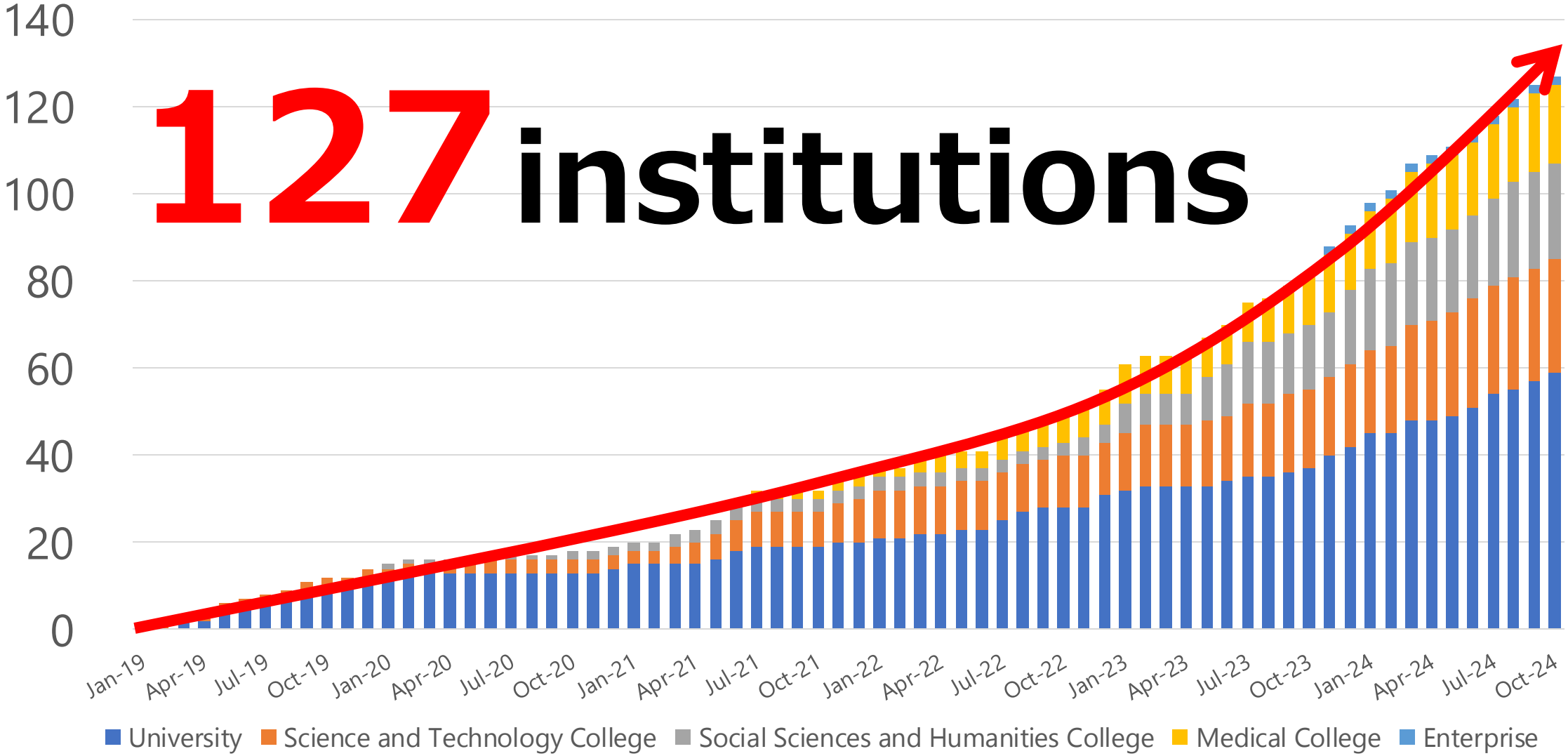


GakuNin RDM is part of NII RDC, a research data infrastructure operated by NII, and specializes in handling closed (private, restricted sharing) data during research process. In April 2021, we started providing RDM service to academic institutions across Japan with an operational system with 24/7 operators. NII adopted Open Science Framework (OSF), developed by the Center for Open Science (COS) in the US, as the essential software for GakuNin RDM. NII has implemented new functions required for Japanese academic institutions.

Concept of GakuNin RDM service



Increase in the number of institutions using GakuNin RDM



NII Research Data Cloud

**Accelerate Open Science and Data Intensive Science
by NII Research Data Cloud with New *Seven* Features**

Reuse

Code Package

Improves reproducibility of research outputs by packaging data, program, and data analysis environment.

Trust

Data Provenance

Provides an incentive model for data publication by monitoring data usage provenance.

Safe

Secure Storage

Preserves sensitive information through the ultra-secure and robust storage environment.

Secure Storage Environment

Management

Data Governance

Transforms DMP into a crucial project document by supporting machine-actionable functions.

Interoperable

Data Curation

Contributes data reuse by developing an eco-system for data curation networks.

Protection

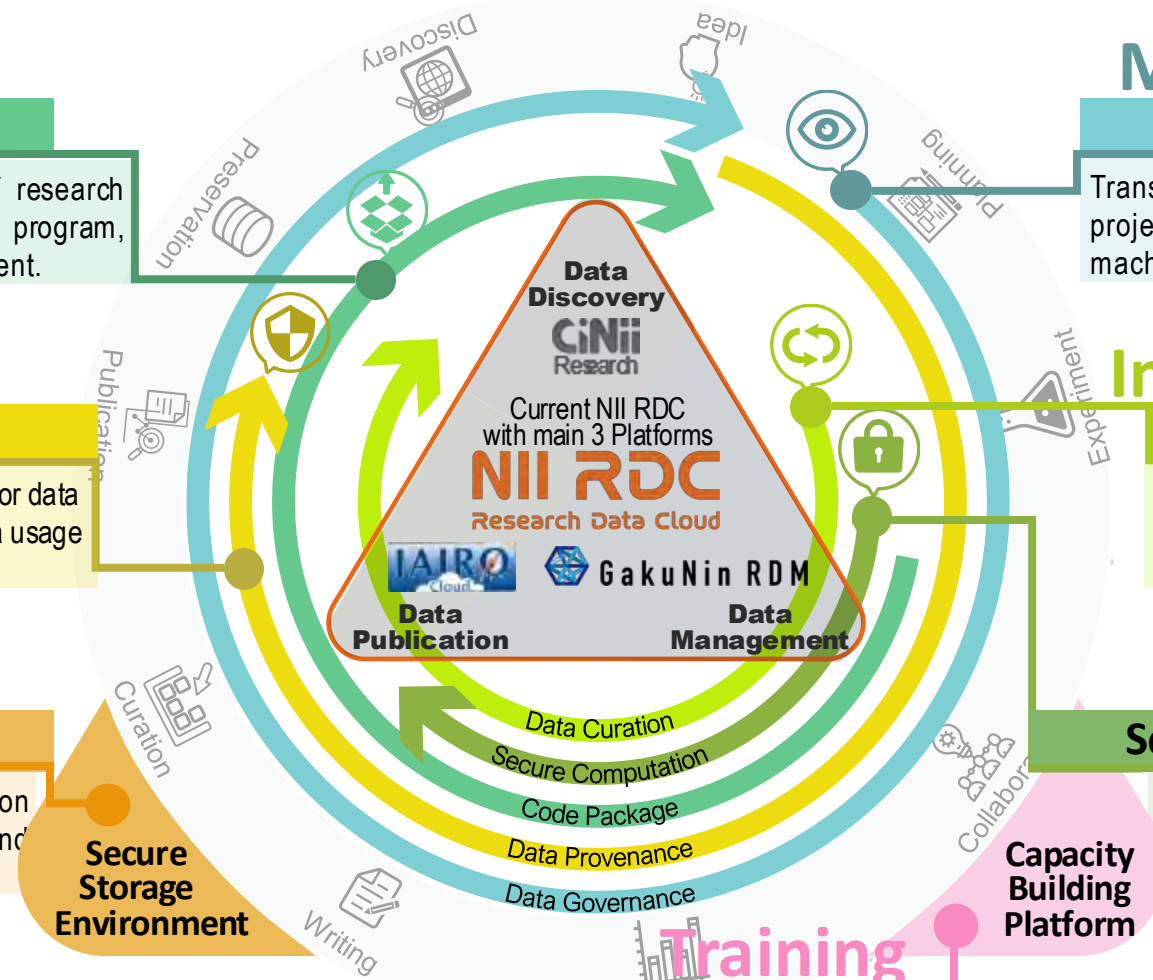
Secure Computation

Creates super-secured data analysis space by providing secure computing technology.

Capacity Building Platform

Capacity Building

Trains all researchers to become new science practitioners by providing a learning environment for RDM skills.



Walkthrough of Research Data Management

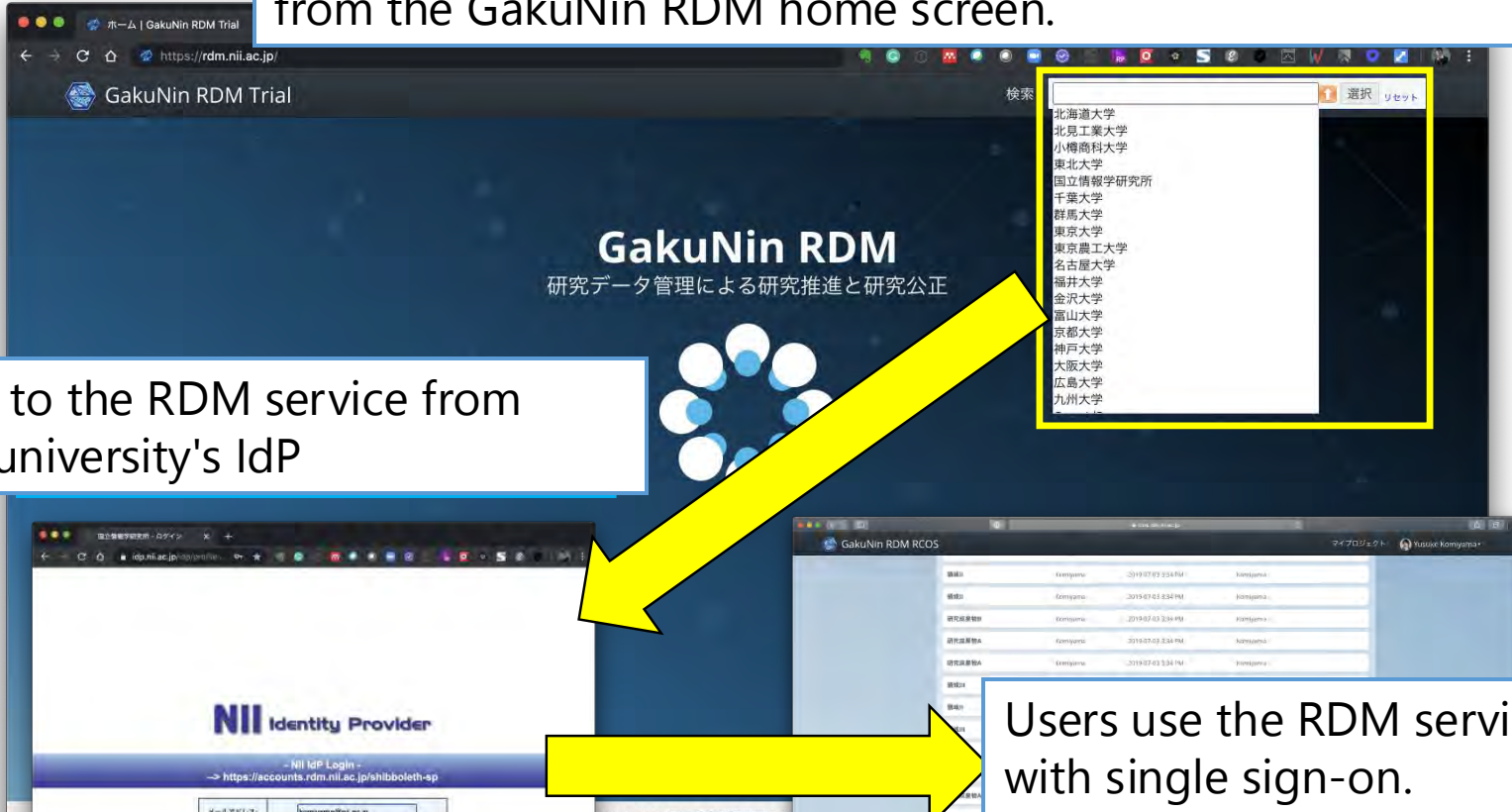
(1) Research-data sharing function with AAI

User selects their institution in the discovery service from the GakuNin RDM home screen.

Login to the RDM service from each university's IdP

Users use the RDM service with single sign-on.

GakuNin RDM is registered as a service provider in the GakuNin federation (AAI: authentication and authorization infrastructure), and users can use the system only by linking with institutional identity providers participating in GakuNin.



(1) Research-data sharing function with AAI

Specializing in the management of undisclosed research data

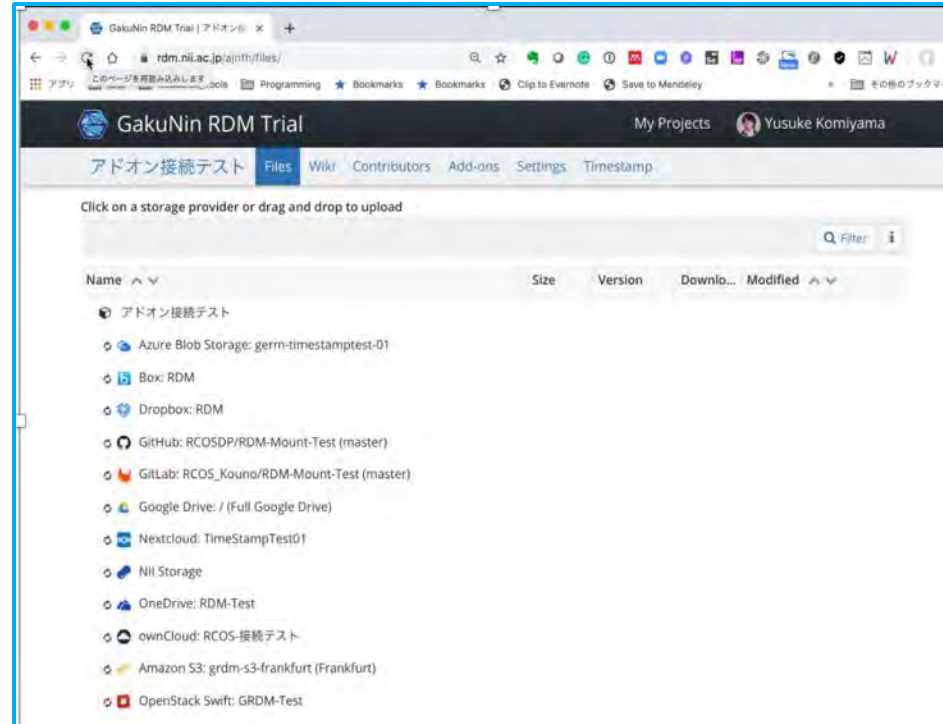
The screenshot shows the GakuNin RDM Trial interface for the project 'NII-OF2019'. The interface includes a header with the project name and a navigation menu. The main content area is divided into several sections:

- Title:** Points to the project name 'NII-OF2019'.
- Contributors:** Points to the list of contributors: Yusuke Komiyama, Hidetoshi Yoshimoto, 平原 泰明, and Mitakazu Hayashi.
- Date:** Points to the creation date '2019-05-26 10:47 PM' and the last update date '2020-06-05 10:53 AM'.
- License:** Points to the license information 'CC BY Attribution 4.0 International'.
- Abstract:** Points to the project description: '2019年度のNII学術情報基盤オープンフォーラム用のプロジェクト'.
- File Manager:** Points to the file list table, which includes columns for name, version, and last update time. Files listed include 'NII-OF2019_AXIES-NII_Panel.pptx', 'NII-OF2019_Contents-OS1_track_All.pdf', and 'NII-OF2019_Contents-OS1_track_All.s.pdf'.
- Citation:** Points to the '引用' (Citation) section.
- Related Projects:** Points to the 'コンポーネント' (Component) section.
- Tags:** Points to the 'タグ' (Tag) section, which shows tags like 'オープンフォーラム', 'GakuNin RDM', 'GRDM', and 'DF2019'.
- Recent Activity:** Points to the '最近の活動' (Recent Activity) section, which shows a log of user actions such as 'Yusuke Komiyama updated wiki page Home to version 1 of NII-OF2019'.

(2) Linking NII research data infrastructure with external tools

Connecting to various cloud services

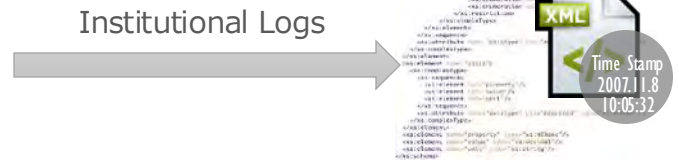
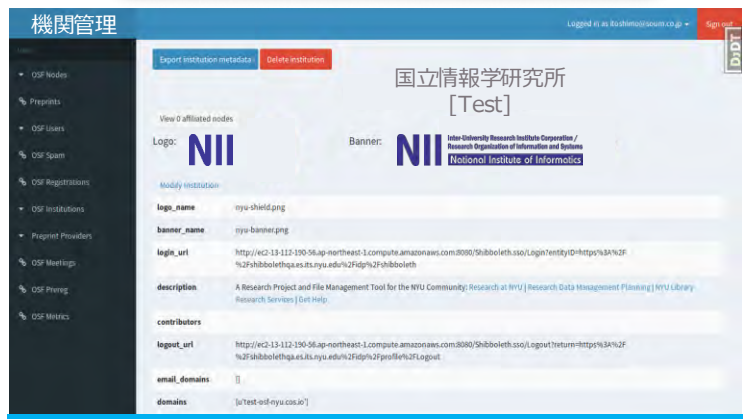
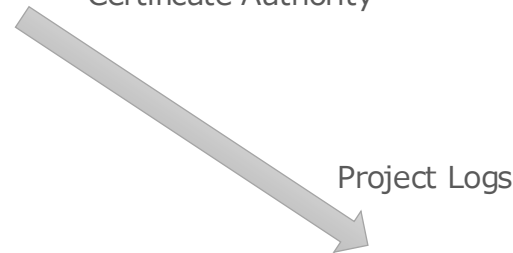
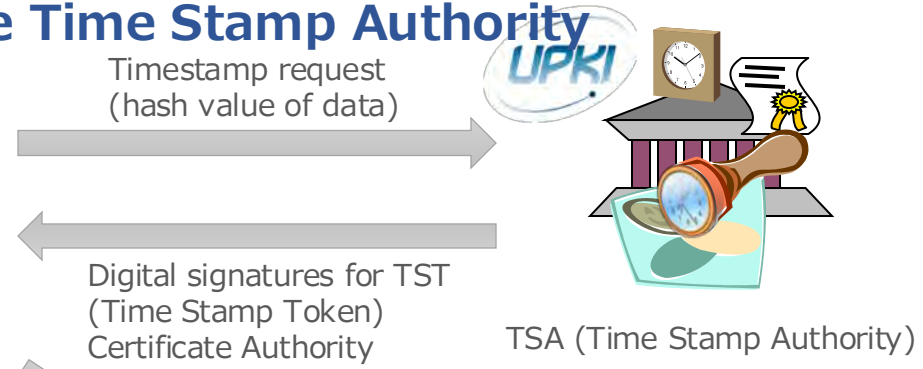
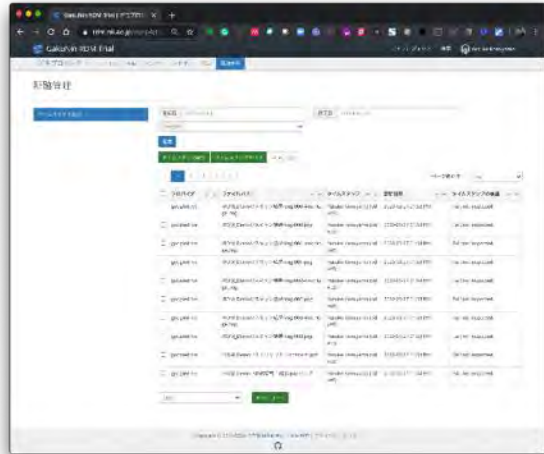
- **Cloud Storage**
 - **Public Cloud**
 - Amazon S3
 - Box
 - Dropbox (personal)
 - **Dropbox for Business**
 - Google Drive
 - One Drive
 - **Azure Blob Storage**
 - **S3 compatible storage** (Wasabi, Sakura Cloud, IDCF)
 - **Private Cloud**
 - ownCloud
 - **Nextcloud**
 - **OpenStack Swift (API ver2, ver3)**
 - **S3 compatible storage (on-premise)**
- **Figure and Tables Sharing**
 - Figshare
- **Source Code Repository**
 - GitHub
 - Bitbucket
 - Gitlab
- **Data Repository**
 - Dataverse
 - **WEKO (NII Publication Platform)**
- **Reference Manager**
 - Mendeley
 - Zotero



Various cloud services can be connected to GakuNin RDM according to the scales and purpose of the academic institution. In red are plug-ins that have been extended by NII, especially from open source.

(3) Preservation of research trails

Proof of Data Content by the Time Stamp Authority



Detection of data manipulation and content changes without the use of GakuNin RDM principal investigators and institutional administrators manage research trails in one place.

(4) Administrative functions for institutional use

Supporting system management and customization for agency administrators

The screenshot displays the 'アドオン利用制御' (Add-on Usage Control) page for GakuNin RDM IdP. The interface includes a sidebar with navigation options like 'RDM ノード', 'プレプリント', and 'RDM ユーザ'. The main content area shows a list of add-ons with checkboxes for enabling them: Amazon S3, Azure Blob Storage, Bitbucket, Box, Dataverse, Dropbox, and Dropbox Business / チームメンバーファイルア. Below this, there are two line graphs: '登録ファイル数' (Registered File Count) and 'ファイルの合計サイズ' (Total File Size), both showing an upward trend from February 2017 to December 2017. A table above the graphs provides monthly data for these metrics.

Feb-2017	Mar-2017	Apr-2017	May-2017	Jun-2017	Jul-2017	Aug-2017	Sep-2017	Oct-2017	Nov-2017	Dec-2017
88	134	142	145	167	203	216	226	234	250	273
85	106	121	124	154	178	187	206	209	243	254
100	200	200	200	260	290	300	300	360	500	500
65%	24%	24%	25%	31%	38%	37%	42%	42%	45%	51%

- Connect University-owned storage
- Selection of available external services
- Logging of Institutional Users

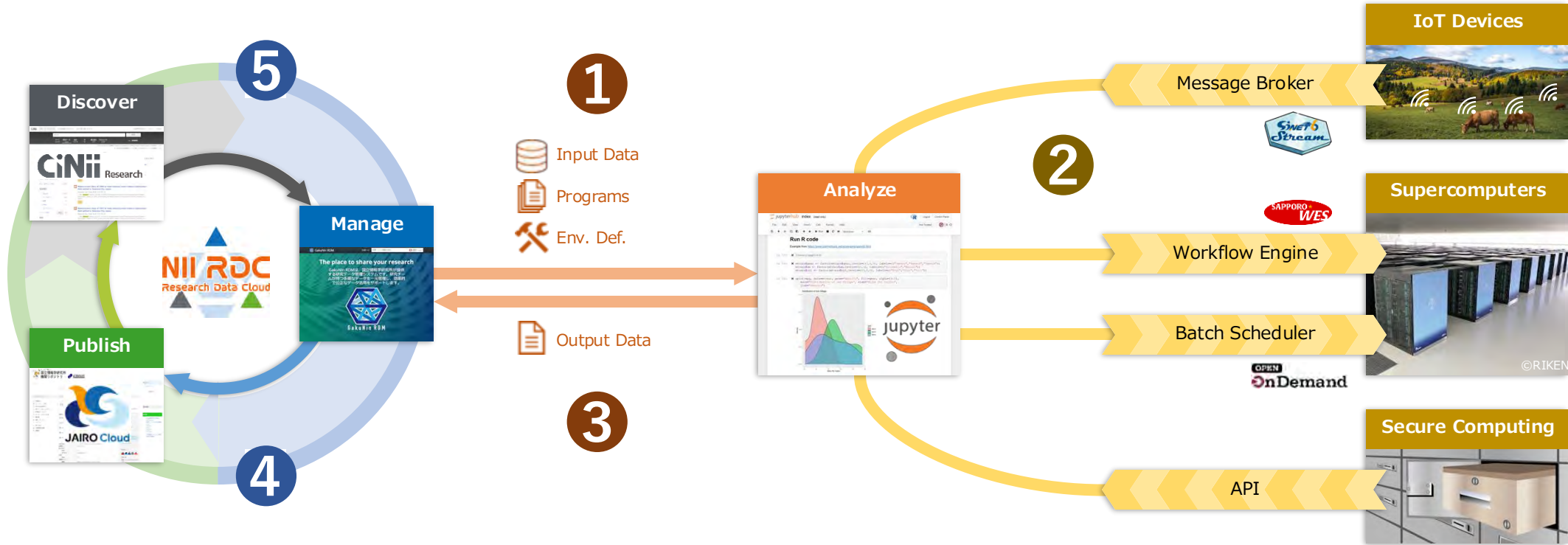
Advanced Applications: Code Package Function

(Data Analysis Function)

Background of Code package Function

- To promote data-driven science, a high degree of integration between research data and computational resources is desired. .
- NII is developing a variety of computing services that cooperate with NII RDC to facilitate the use of research data by diverse researchers.

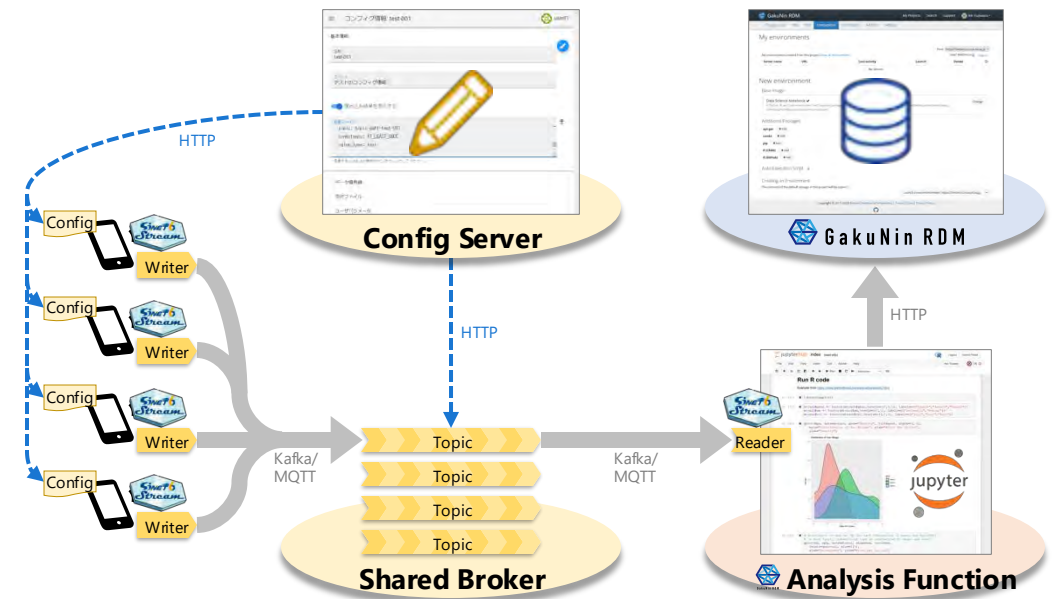
Code package Function



- ① Auto-build a JupyterHub environment on the NII Cloud from within GakuNin RDM
- ② Cooperate with a variety of data sources and computational resources
- ③ Share the analysis results with the project members in GakuNin RDM
- ④ Publish the project as a "computational reproduction package"
- ⑤ Discovered and reused by other researchers to start a subsequent research

Systems Development of Code Package Function

- The SINETStream Trial Service provides a framework for researchers to collect and analyze streaming data without having to deploy a broker.
- The GakuNin RDM Data Analysis Function provides researchers with a Jupyter environment on NII Cloud. Researchers can run Python, R, and MATLAB programs to analyze their data stored in GakuNin RDM.
- The Reproducible Package Function archives a GakuNin RDM project in an RO-Crate format and makes it available for reuse by other researchers through JAIRO Cloud and CiNii Research.
- The GakuNin RDM App for Open OnDemand mounts a GakuNin RDM's storage from within a supercomputer. It is currently supported at Fugaku at RIKEN.



Advanced Applications: Data Governance Function

(Data management Plan) Function

Status of DMP (Data Management Plan) use

It is difficult for researchers to utilize DMPs required by funding agencies (FAs) in their research activities.

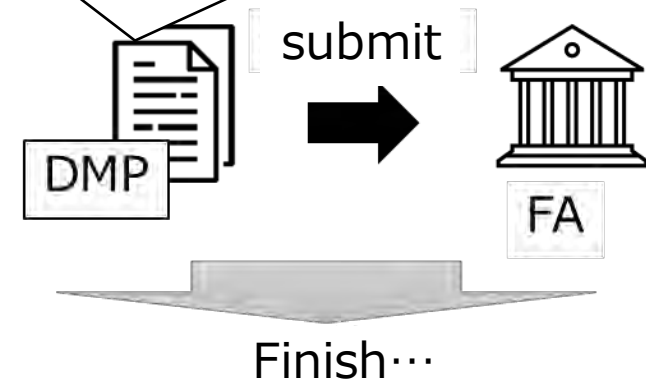
- ✓ DMPs are created as free-form text, which is difficult for a computation resource to interpret.
- ✓ RDM according to a DMP is left to researchers.
- ✓ It's troublesome for both researchers and institute staff to verify the consistency between DMPs and the state of RDM.



DMPs are treated as just documentations...

An example of a DMP

No.	Data Name	Description	Manager	Class	Publication level	Reason secrecy
1	Data captured by sensors during the ○ trial, including related data	This includes road image data captured by sensors during the ○ trial.	○○ Research Institute	Commi ssioned Data	Level 4 (Broad provision and utilization))	Not confidential

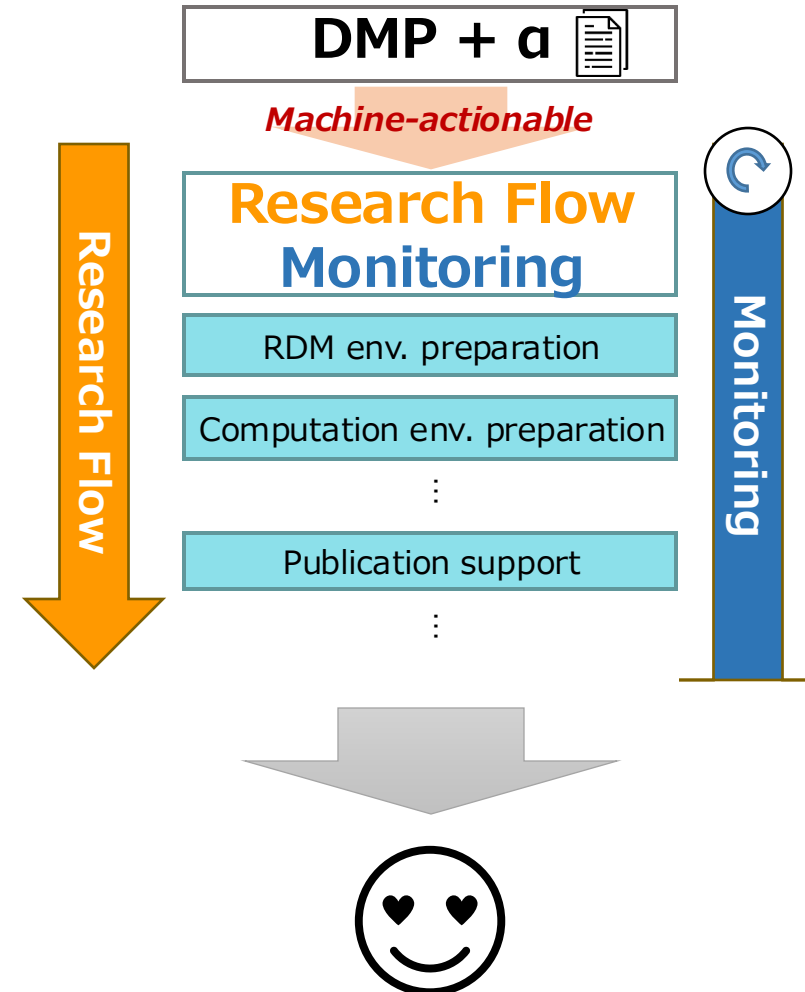


Solutions

We have been developing functions that **semi-automatically** generate an appropriate environment for RDM according to a DMP + α (e.g., rules defined by researchers).

- ✓ **Research Flow:** arrange the research procedure according to DMP + α .
- ✓ **Monitoring:** verify the consistency between DMP + α and the state of RDM continuously.

↓
Keep/improve the quality of RDM!



Vision of Data Governance Functions

Issue: More and more funding agencies request researchers DMPs

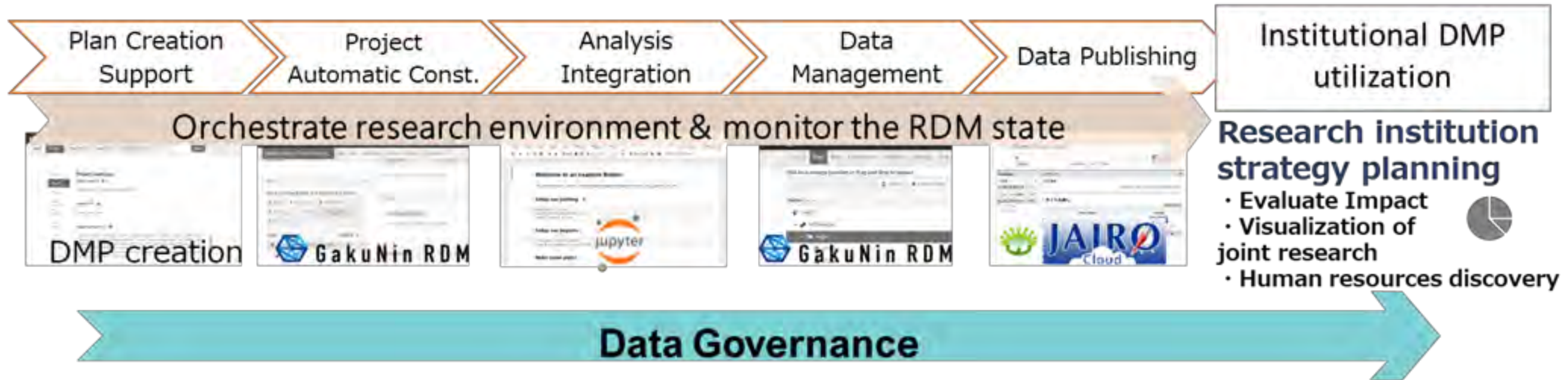
Additional work to do?



Why should I submit?



Solutions: **Automation of data management process based on DMP**

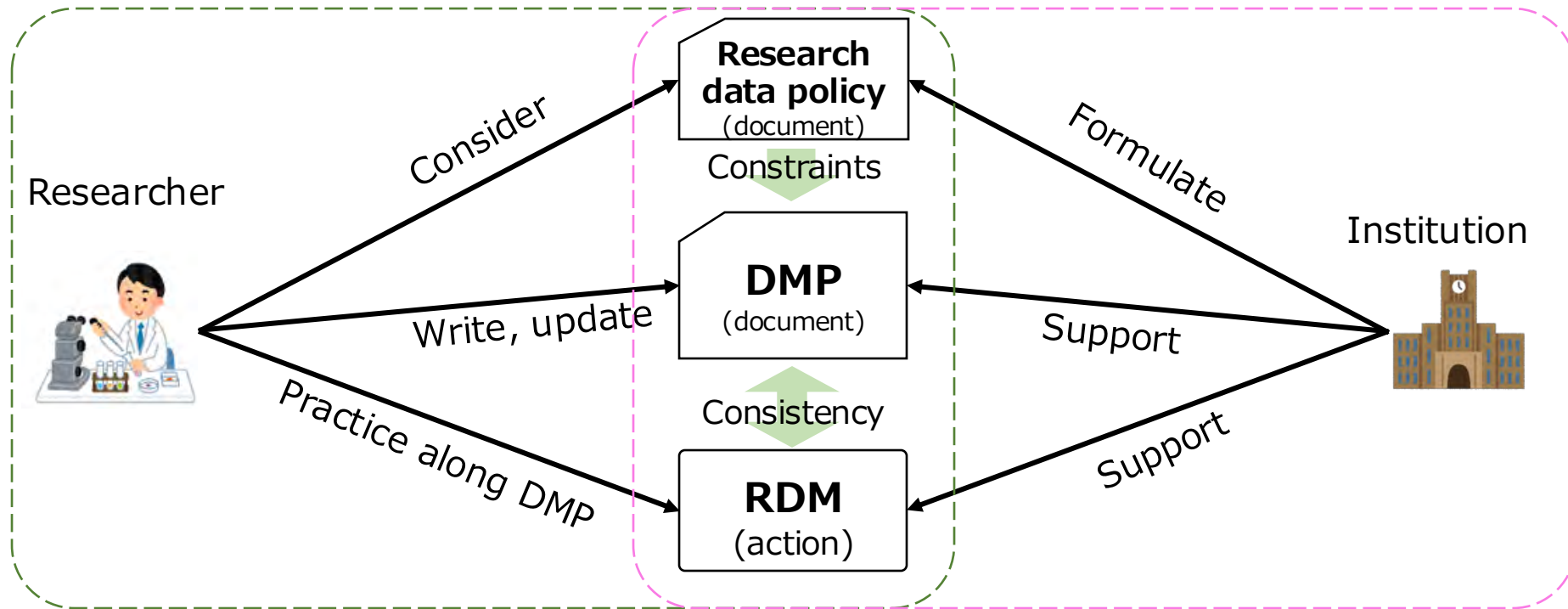


The data governance function builds an environment in accordance with DMPs and policies and monitors whether data management is being conducted in compliance with the DMP.

While paying attention to international trends in maDMP, we are advancing its development at NII.

Model of research data governance

Disciplined behavior/actions to protect and improve the quality of research data management (RDM).



As a researcher:

- Consider the research data policy (RDP).
- Practice RDM along their DMP.
- Utilize the environment for RDM (provided by her/his org.).

As an institution:

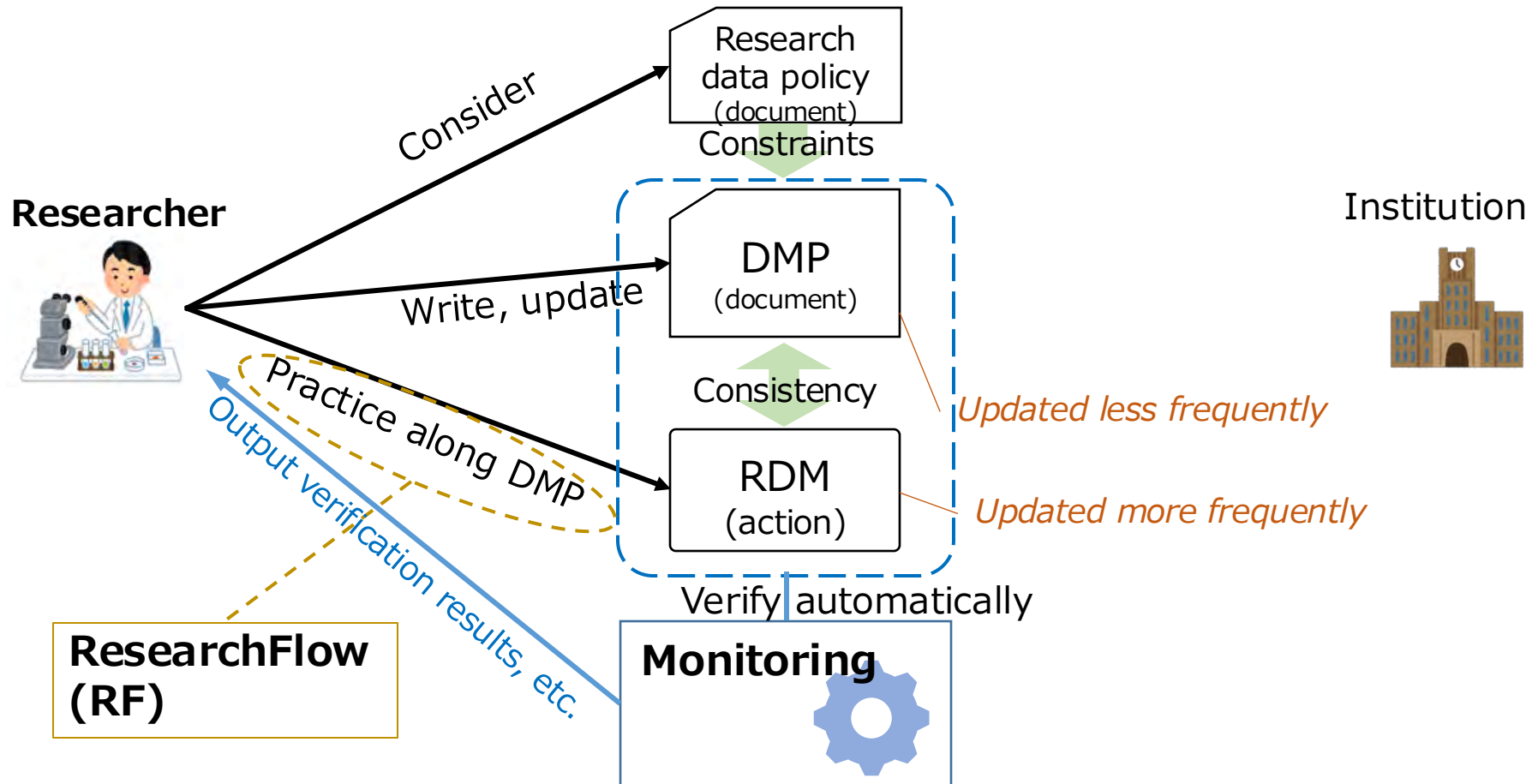
- Formulate and maintain their RDP.
- Practice DG along its RDP.
- Support their researchers to practice RDM.

Our first concept of DG functions

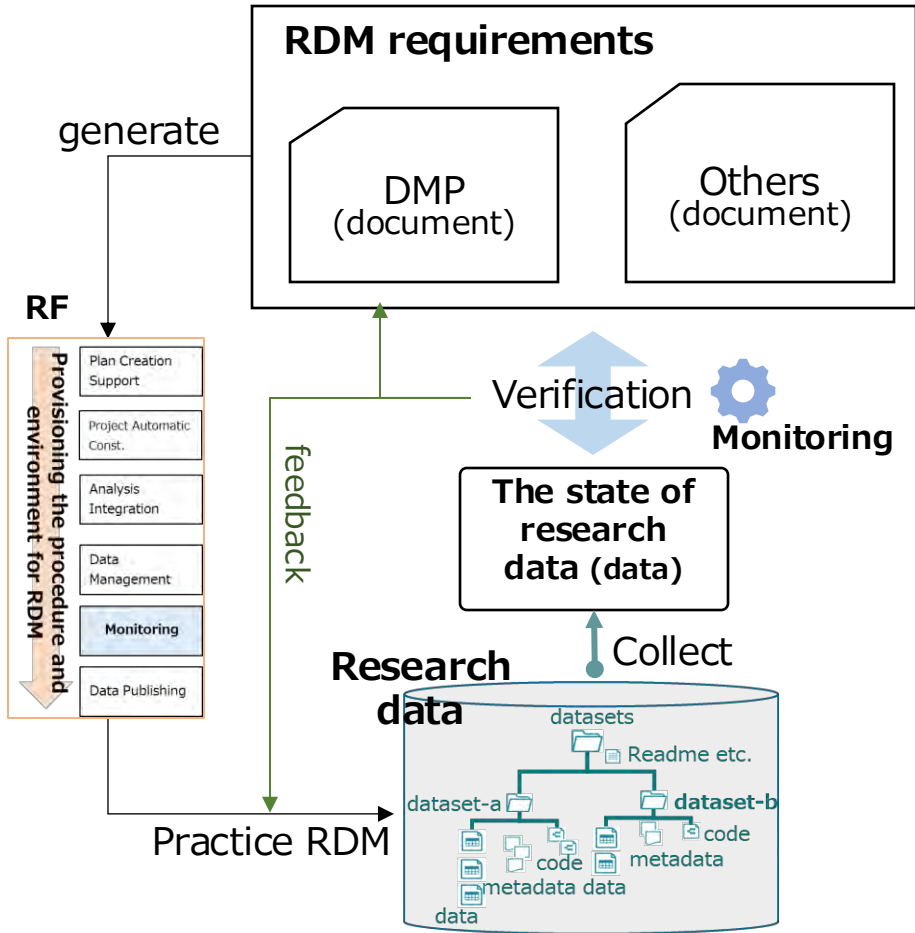
DG as a researcher

Support a part of DG as a researcher:

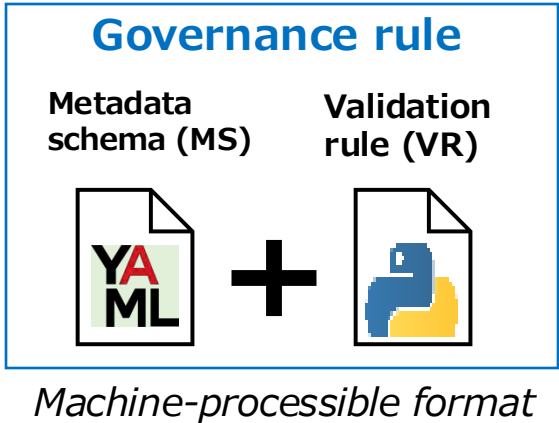
- Verify whether the state (metadata + dataset) of research data satisfies the RDM requirements described in the DMP.
- Support researchers to practice RDM along their DMP.



How to apply RDM requirements



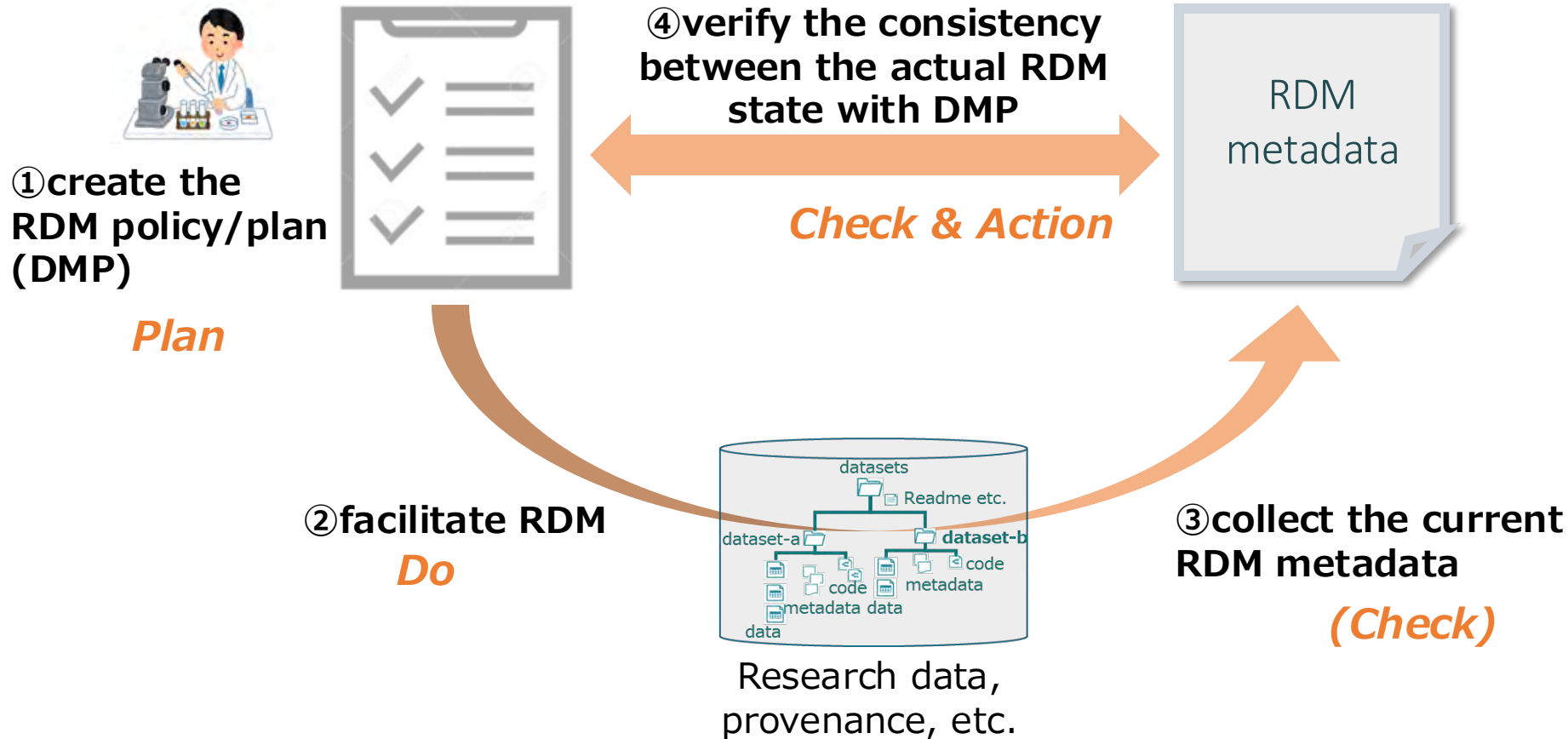
- In terms of verification:
- criteria, how to judge (for verification).
- In terms of provisioning:
- requirements of the env.



Fundamental DG functions

What we provide: A framework for implementing the PDCA cycle for RDM

Value: Visualization of data management policy and management status, labor saving of data management



Summary

- The National Institute of Informatics (NII) in Japan is advancing open science through its NII Research Data Cloud (NII RDC), designed for academic institutions.
- Within NII RDC, GakuNin RDM (GRDM) serves as a platform for managing and sharing controlled data. GRDM is based on the Open Science Framework (OSF), which was developed by the Center for Open Science (COS) in the United States, and has been tailored to meet the specific needs of Japanese researchers. GRDM supports authentication and authorization infrastructure (AAI), cloud service integration, and data provenance management.
- NII RDC also offers data analysis capabilities through JupyterHub, enhancing the reproducibility of research.
- Additionally, NII is developing machine-actionable Data Management Plans (maDMPs) to support data governance and ensure compliance with policies and regulations by dynamically defining workflows and assessing data quality.
- This presentation has aimed to provide you with insights into the latest developments in research data management and infrastructure in Japan.

RCOS: Research Center for Open Science and Data Platform



RCOS

rcos-ext@nii.ac.jp