

Design of Research Data Governance Functions Using Data Management Plans and Records

○Toshiyuki Hiraki, Miho Funamori, Shigetoshi Yokoyama,
Yusuke Komiyama, Mikiko Tanifuji, Kazutsuna Yamaji

National Institute of Informatics, Japan

Table of Contents

1. Scope and issues of Institutional Research Data Governance (IRDG)
2. IRDG components
3. IRDG tool & its components
4. Usage scenarios of IRDG tool
5. Challenges in IRDG tool

1. **Scope** and **issues** of **Institutional Research Data Governance (IRDG)**

Scope of Institutional Research Data Governance (IRDG)

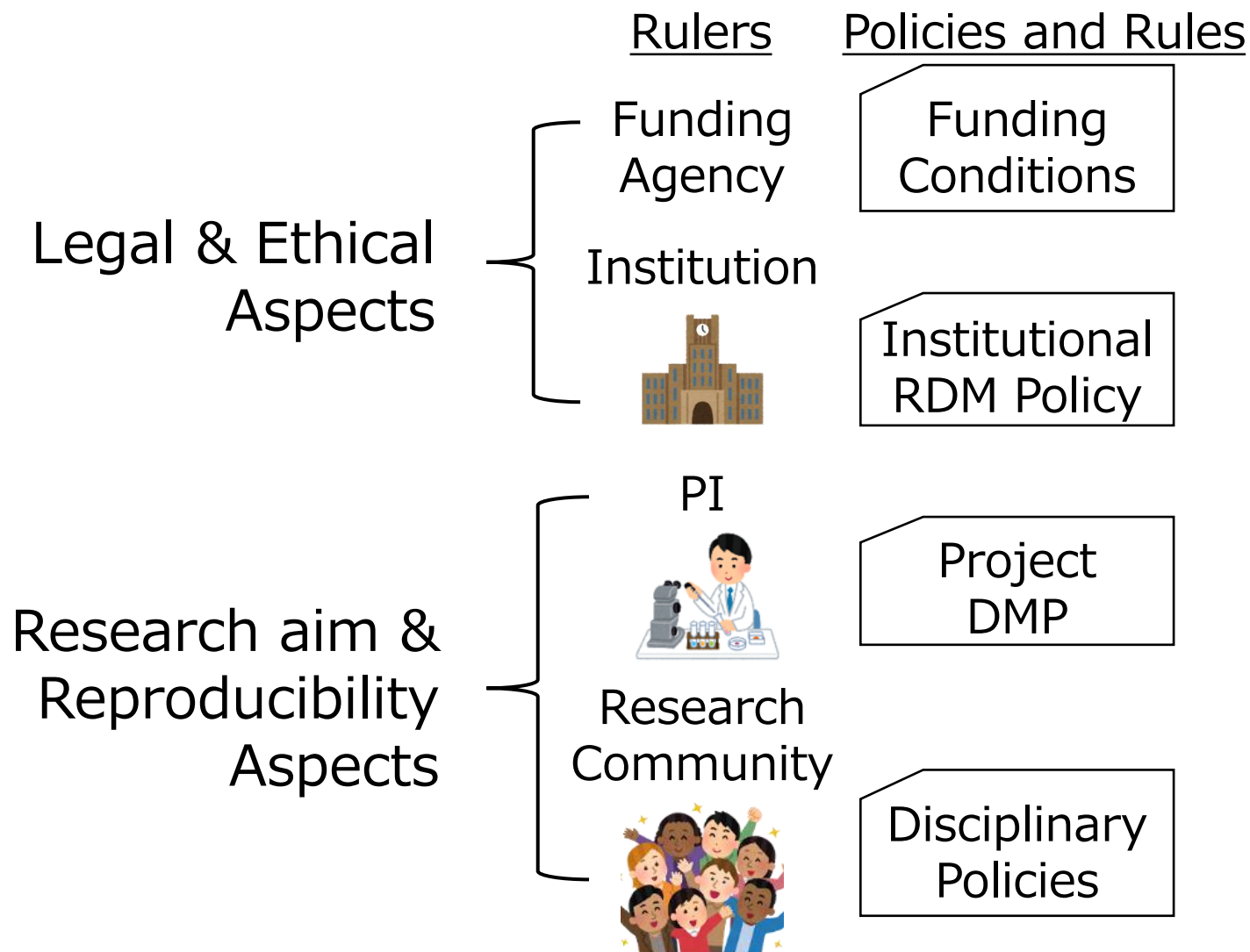
[Why IRDG?]

- Institutions need to manage data to ensure research compliance and research integrity.

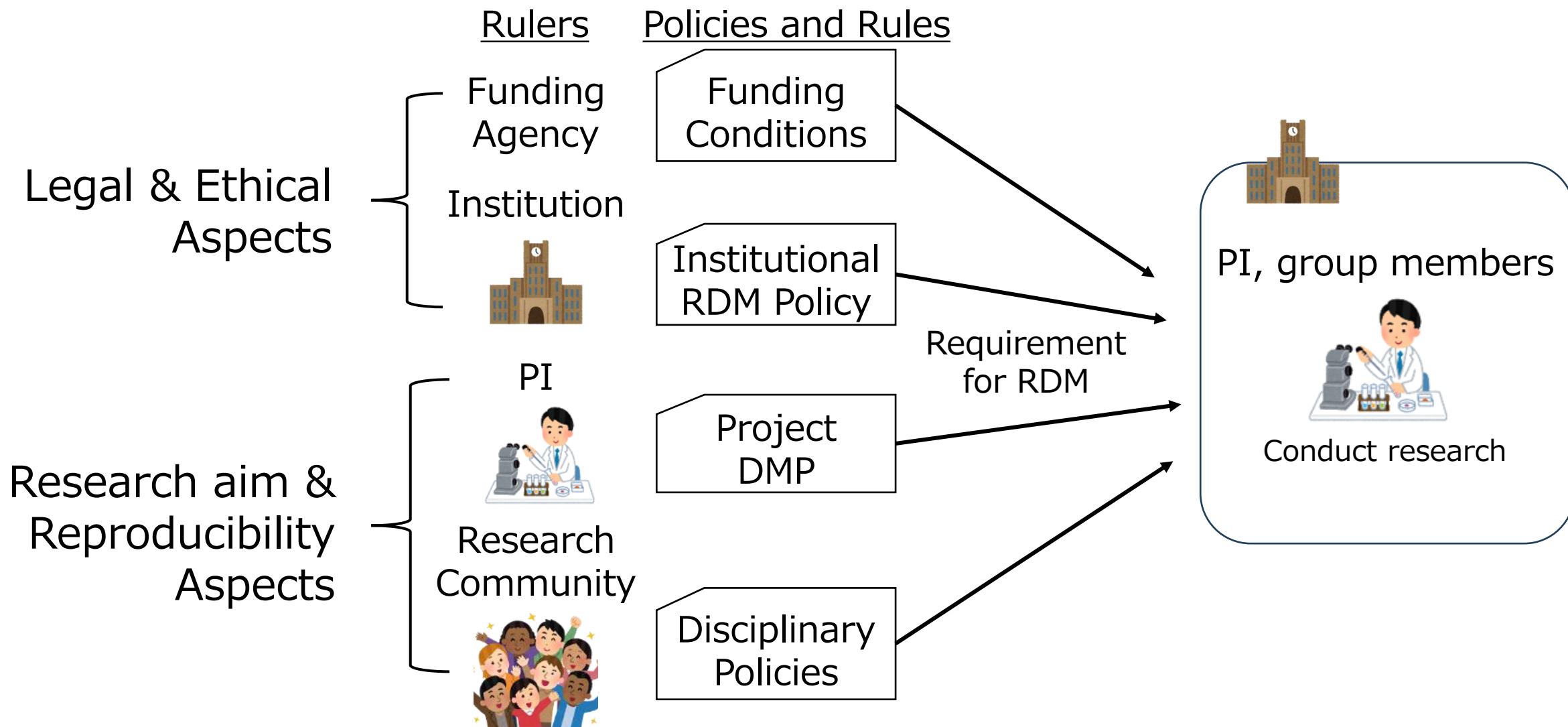
[What should they do?]

- Institutions and researchers are required to realize IRDG in accordance with the policies relevant to research data management (RDM).

RDM under policies and rules

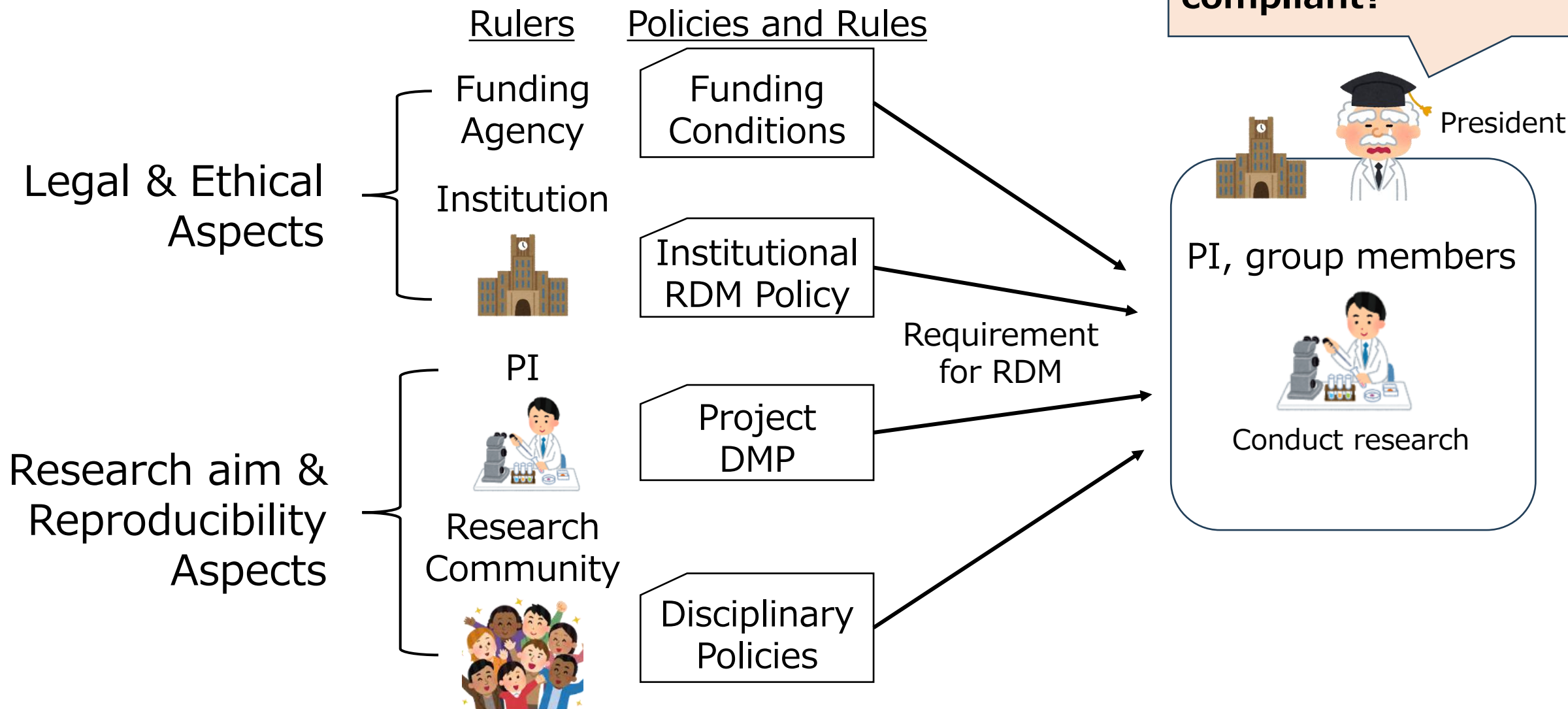


RDM under policies and rules



RDM under policies and rules

How can I make RDM by researchers compliant?



Issues: How can institutions make RDM by researchers compliant?

1. Institutions need to track RDM by researchers.
2. Institutions need to make sure that researchers' RDM is following RDM policies.
3. Institutions and researchers cannot take on the heavy burden of RDM.

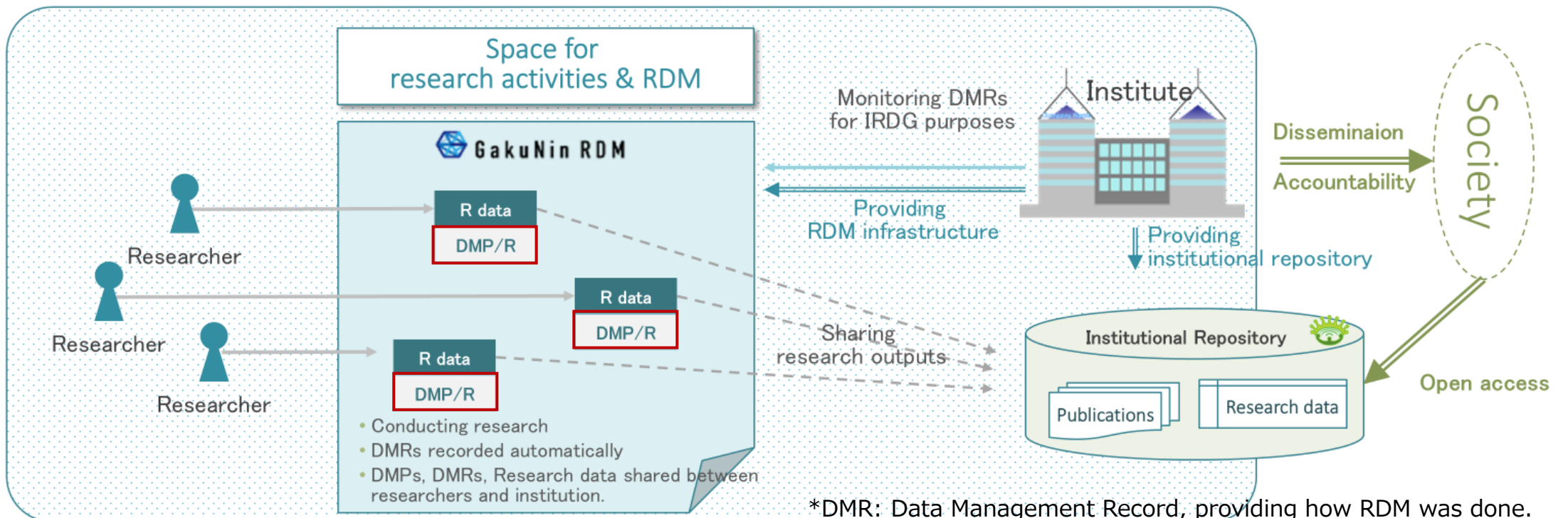
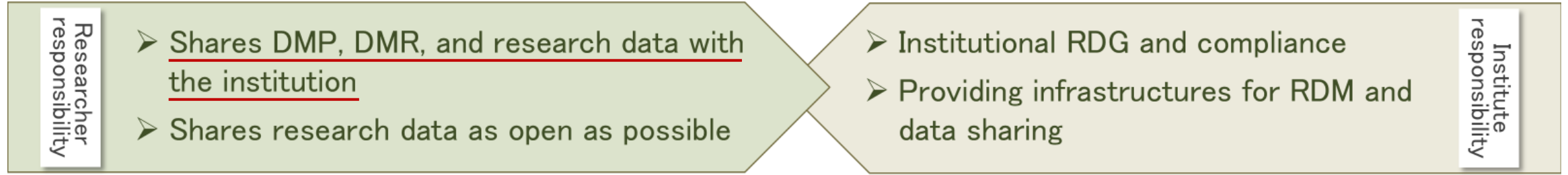
Issues: How can institutions make RDM by researchers compliant?

1. Institutions need to track RDM by researchers. → **RDM recording**
2. Institutions need to make sure that researchers' RDM is following RDM policies. → **RDM checking and adjustment**
3. Institutions and researchers cannot take on the heavy burden of RDM. → **IRGD tool**

2. IRDG components

RDM recording system

Institutions can track research data management by researchers if they provide RDM services with shared storage.



*DMR: Data Management Record, providing how RDM was done.

What should institutions track?

Institutions need to know if researchers' RDM was conducted in accordance with policies.

→ Need to record RDM for each policy item.

→ Realized by DMP^{*1} + DMR^{*2}.

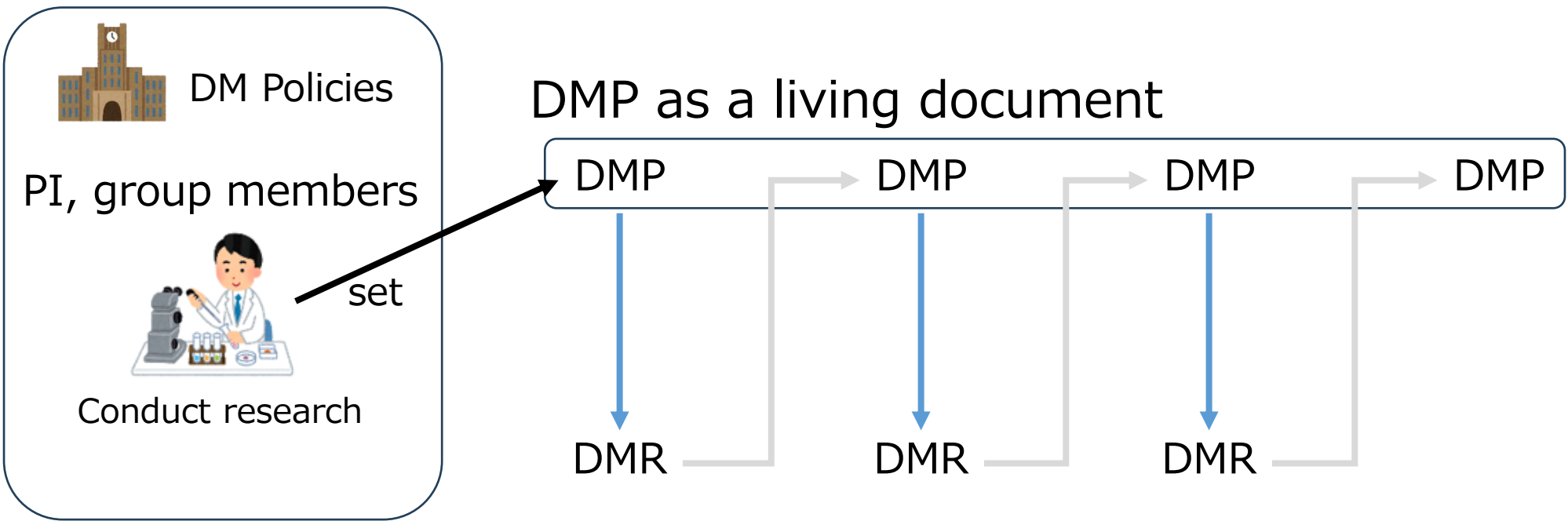
*1 DMP: Data Management Plan

*2 DMR: Data Management Record, providing how RDM was done.

Items of DMP/DMR

- Project
- People
- Ethics and Security
 - Information Security Classification (public/internal/sensitive/confidential)
 - Research involves:
 - ✓ *Human participant data*
 - ✓ *Use of animals*
 - ⋮
 - Ethics approval needed (y/n)
- Data Collection and Storage
 - Data collection methodology
 - File type, file format
 - Storage location
- Data Retention and Disposal
 - Minimum retention period
 - Data steward
 - Commitments to destroy data prior to end of retention period
- Access and Rights
 - Copyright and intellectual property owners of data created in project
 - Access after the project
 - Use of secondary or third-party data
 - Attach ethics approval, licenses
- Research Workspace

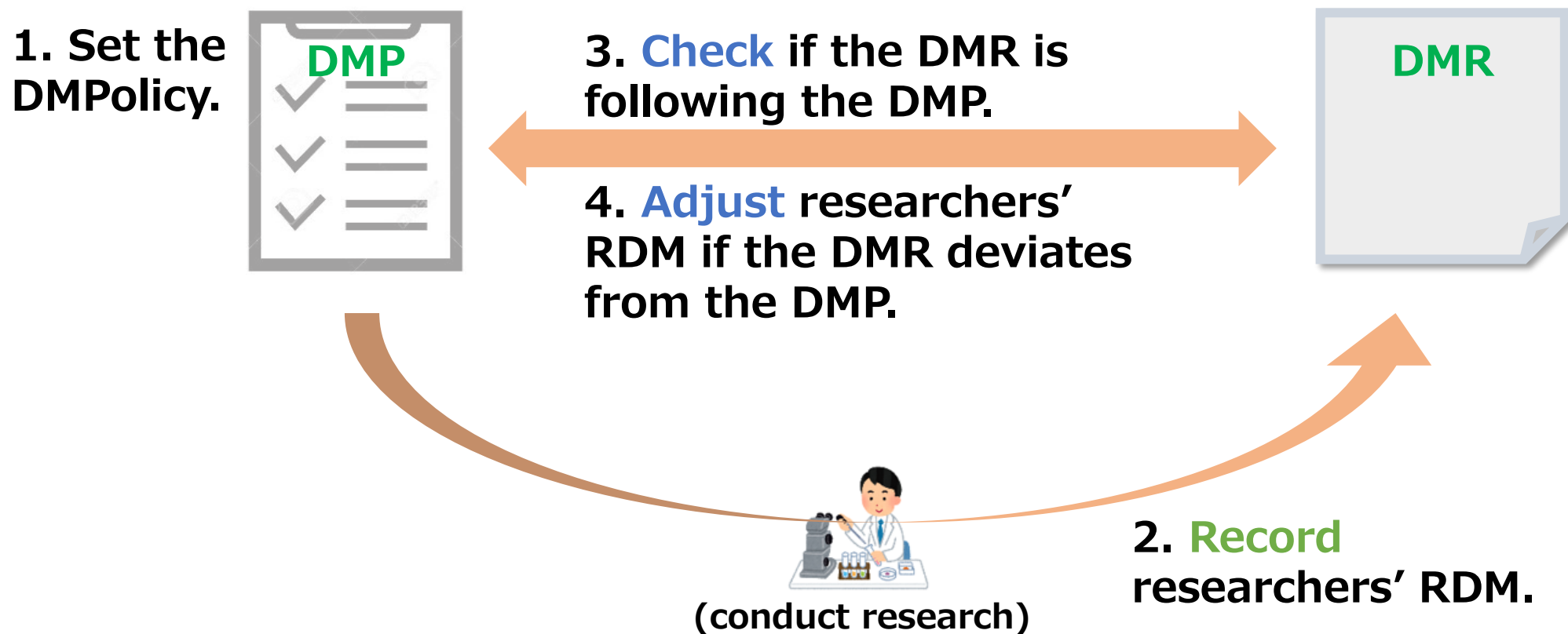
DMP as a living document = DMPolicy + DMRecord



How can Institutions make sure that researchers' RDM is following RDM policies?

1. Set the DMP.
2. Record researchers' RDM (=DMR).
3. Check if the DMR is following the DMP.
4. Adjust researchers' RDM if the DMR deviates from the DMP.

RDM recording, checking and adjustment



The rest issue is ...

1. Institutions need to track RDM by researchers. → RDM recording
2. Institutions need to make sure that researchers' RDM is following RDM policies. → RDM checking and adjustment
3. Institutions and researchers cannot take on the heavy burden of RDM. → **IRGD tool**

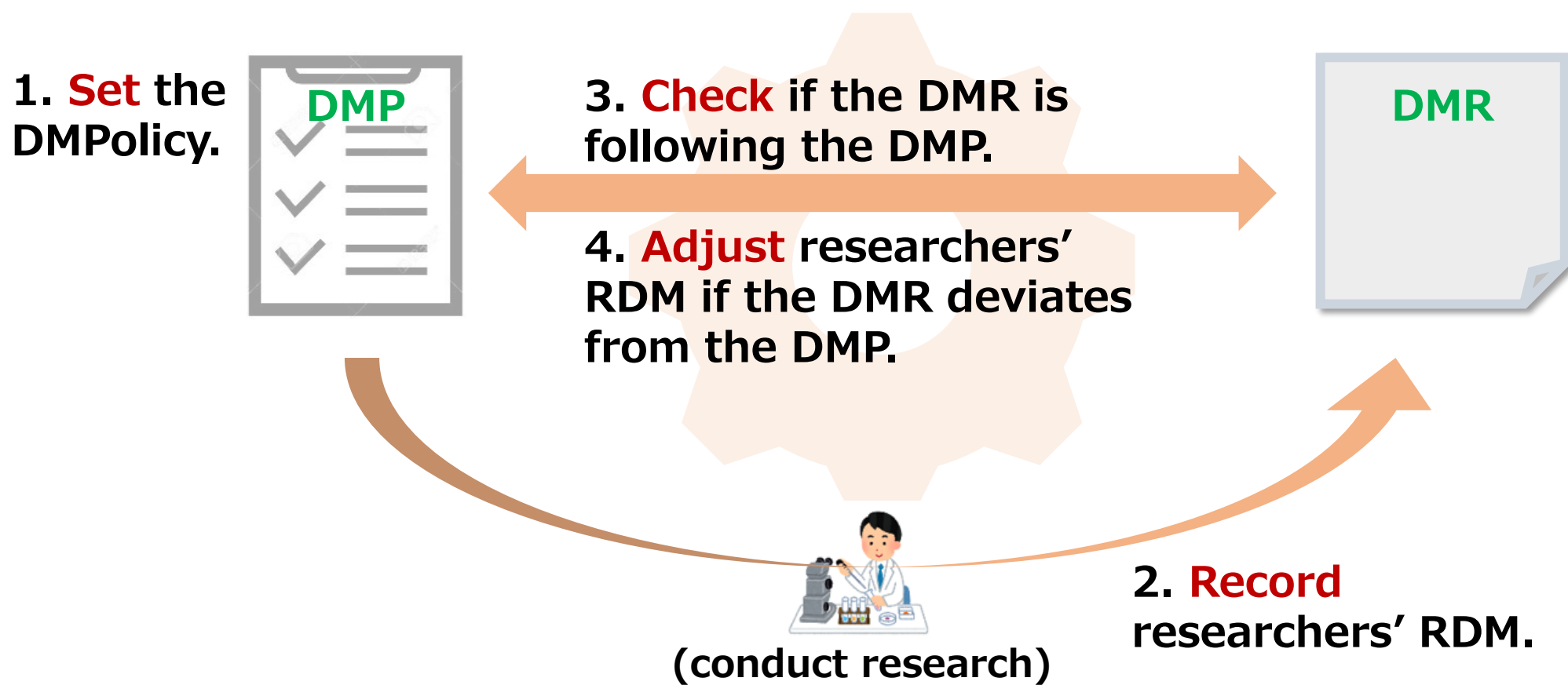
3. IRDG tool & its components

Issues in terms of burden

- It is too burdensome for researchers to report on their research data management to their institutions.
 - Institutions need to make sure that researchers' RDM within the institution are all compliant.
- ➔ We need a tool to assist institutions and researchers in managing RDM by the researchers.

IRDG tool

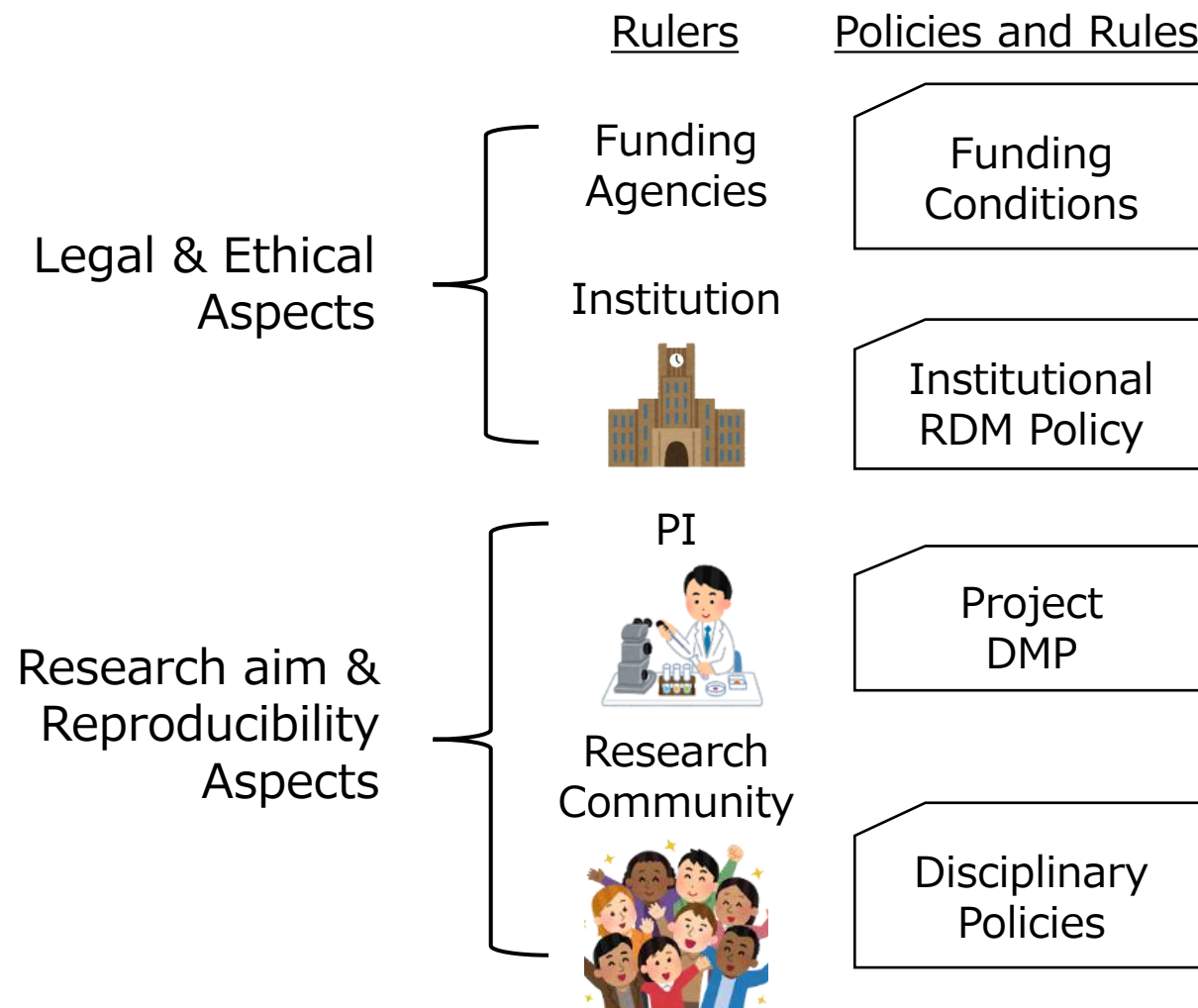
A tool for enabling data governance for ongoing research



IRDG tool

... Component 1: DMPolicy-setting

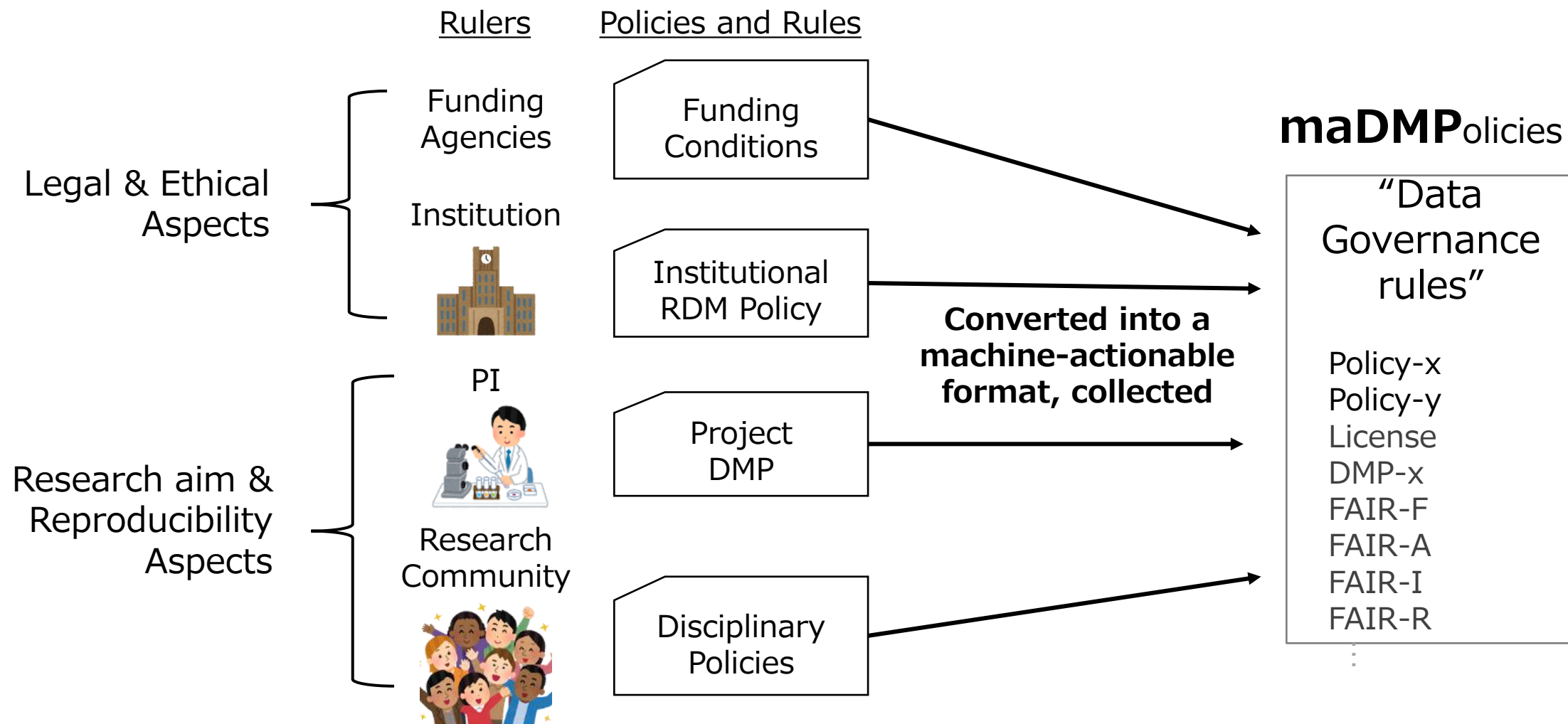
Component to list various RDM policies in a machine-actionable way



IRDG tool

... Component 1: DMPolicy-setting

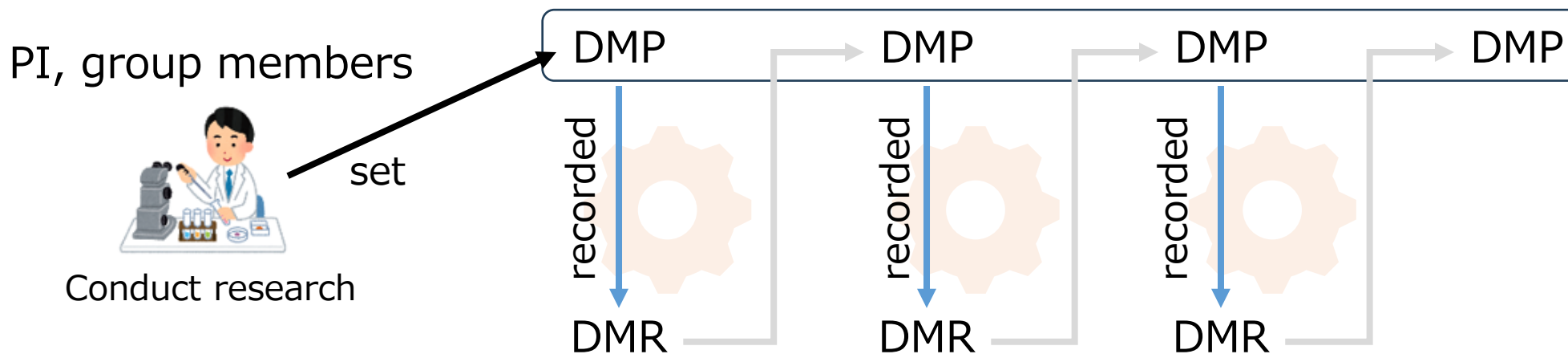
Component to list various RDM policies in a machine-actionable way



IRDG tool

... Component 2: Recording

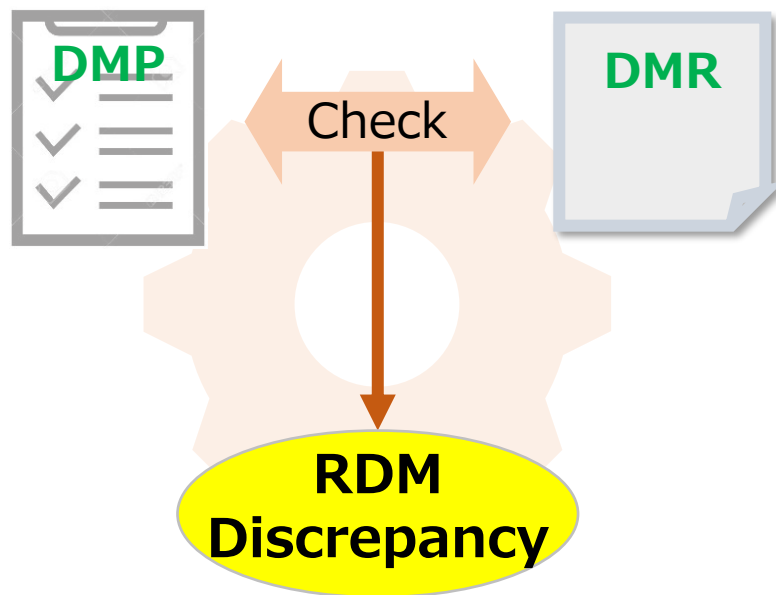
Component to record RDM actions for each DMPolicy item



IRDG tool

... Component 3: Checking

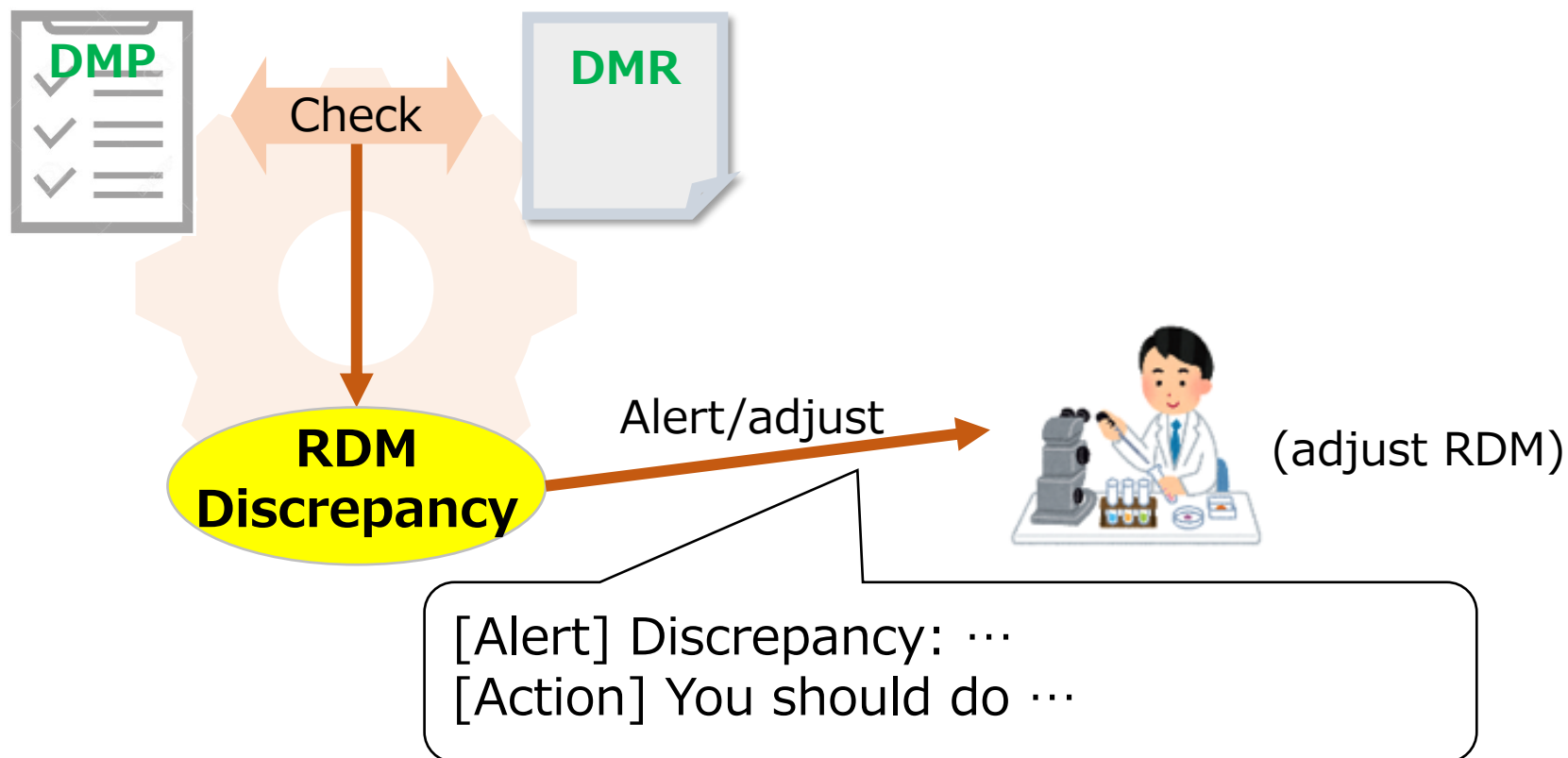
Component to check whether the DMRs satisfy the DMPolicies



IRDG tool

... Component 4: Adjustment

Component to give alerts if there is discrepancy between the DMPolicies and the DMRs



4. IRDG tool usage scenarios

Usage scenario 1: for compliance

Manage research data including personal information

DMP

1. Set the DMPolicy.

[Requirements]

1. Research ethics review required
2. Information consent form required
3. Raw data should stay private
4. Data should be anonymized before sharing



[Alert] **Raw data is public!!**
[Action] Raw data has been made private according to the DMPolicy.

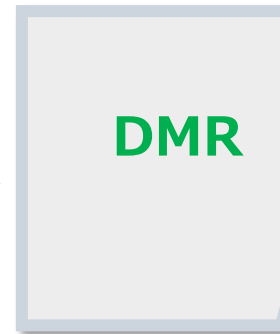
3. Check if the DMR is following the DMP.

4. Adjust researchers' RDM if the DMR deviates from the DMP.

DMR

[RDM actions]

1. Research ethics review passed on 11/15/2023
2. Information consent form obtained on 11/30/2023
3. **Raw data was made public on 2/28/2024**
4. Data was anonymized on 2/28/2024



2. Record researchers' RDM.

(conduct research)

4. Usage scenario 2: for research purpose

Manage research data in accordance with directory structure & file naming rules

DMP

1. Set the DMPolicy.

[Alert] The name of file **20241215_exp1_fin.csv** does not satisfy the file naming rule.
[Action] revise it to satisfy the rule OR the policy.

DMR

[Directory structure]

```
/exp1
  /input
  /output
  /code
```

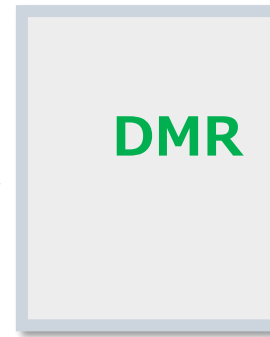
[File naming rule]

In "Output:"
yyyymmdd_exp1_#.csv



3. Check if the DMR is following the DMP.

4. Adjust researchers' RDM if the DMR deviates from the DMP.



[Actual structure]

```
/exp1
  /input
  /xxx.csv
  /output
  /20241215_exp1_1.csv
  /20241215_exp1_2.csv
  /20241215_exp1_fin.csv
  /code
  /main.py
```



(conduct research)

2. Record researchers' RDM.

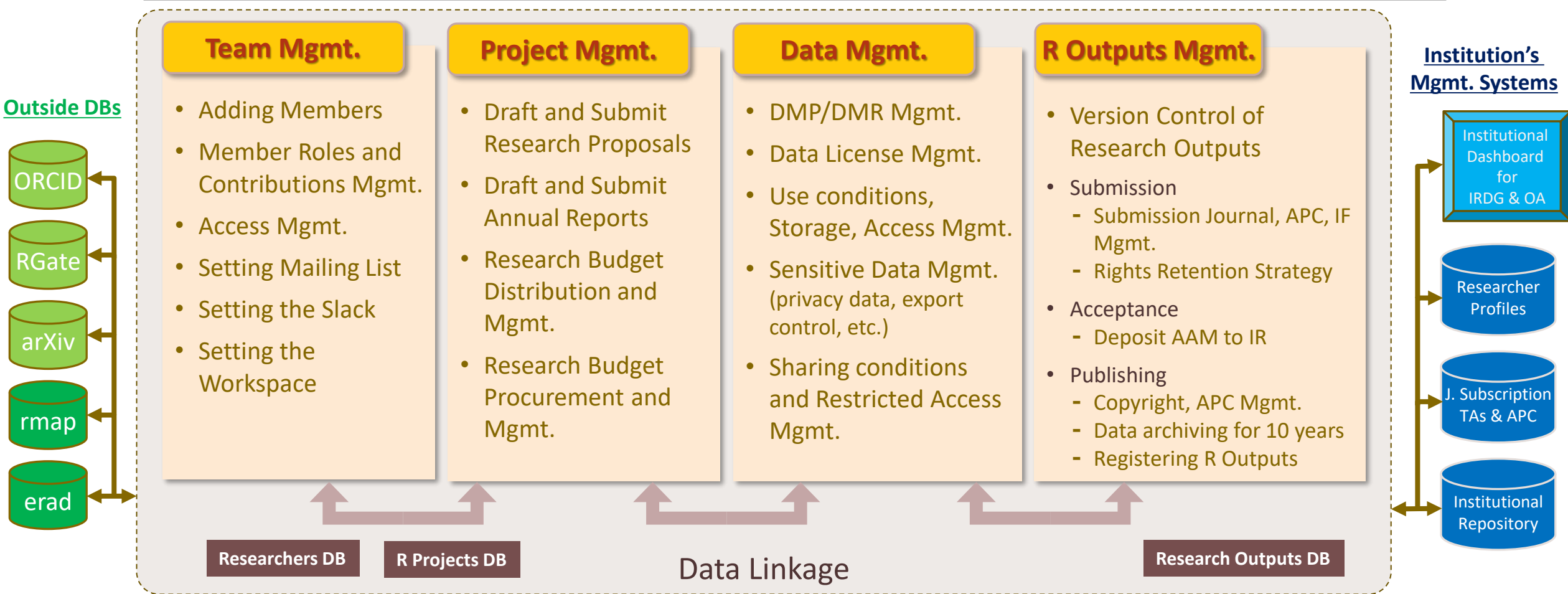
5. Challenges

- Convert **semantic** policies into **machine-actionable** ones.
- Identify **RDM action(s)** for each policy item.

- Clarify **physical RDM workflows** and convert them into **system procedures**.
- Determine the system that the IRDG tool **is installed to**.

- **Change research culture:** set the DMPolicies, share the policies with project members, and conduct RDM along with the policies.

PI's Research Project Management station



Poster #14

A Proposal to use **Data Management Records** for Sustaining Institutional Research Data Governance

Miho Funamori, Toshiyuki Nishiyama Hiraki, Kazutsuna Yamaji
National Institute of Informatics, JAPAN

ABBREVIATIONS

- RDM: Research Data Management
- DMP: Data Management Plan/ Policy
- DMR: Data Management Record.
- IRDG: Institutional Research Data Governance
- NII RDC: NII Research Data Cloud

NII 大学共同利用機関法人 情報・システム研究機構
国立情報学研究所
National Institute of Informatics

