The 2023 eResearch Australasia Conference
Theme: Leveraging Data Assets/Delivering Institutional eResearch Capability
Date and time: 13:40-14:00, Oct. 19th, 2023

NII RCOS
Research Center for Open Science and Data Platform

Research Data Governance Functions of NII Research Data Cloud

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

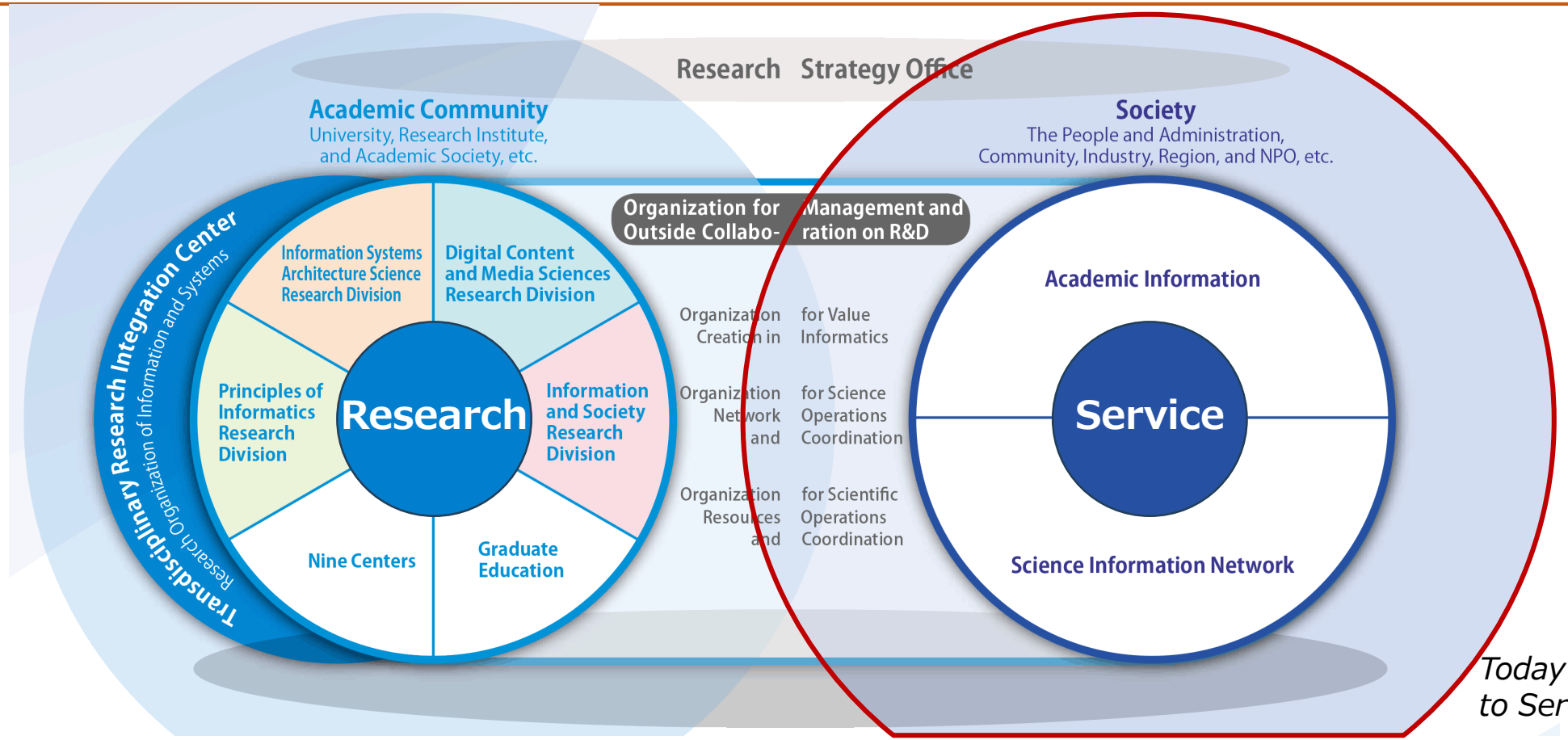
National Institute of Informatics, Japan

© 2023 National Institute of Informatics

Table of contents

1. About us & background
2. Our research data governance (RDG) model
3. Challenges in our RDG model in terms of DG as a researcher
4. Our first solution & vision
5. Summary

About National Institute of Informatics (NII)



Today's talk is related to Service Div.

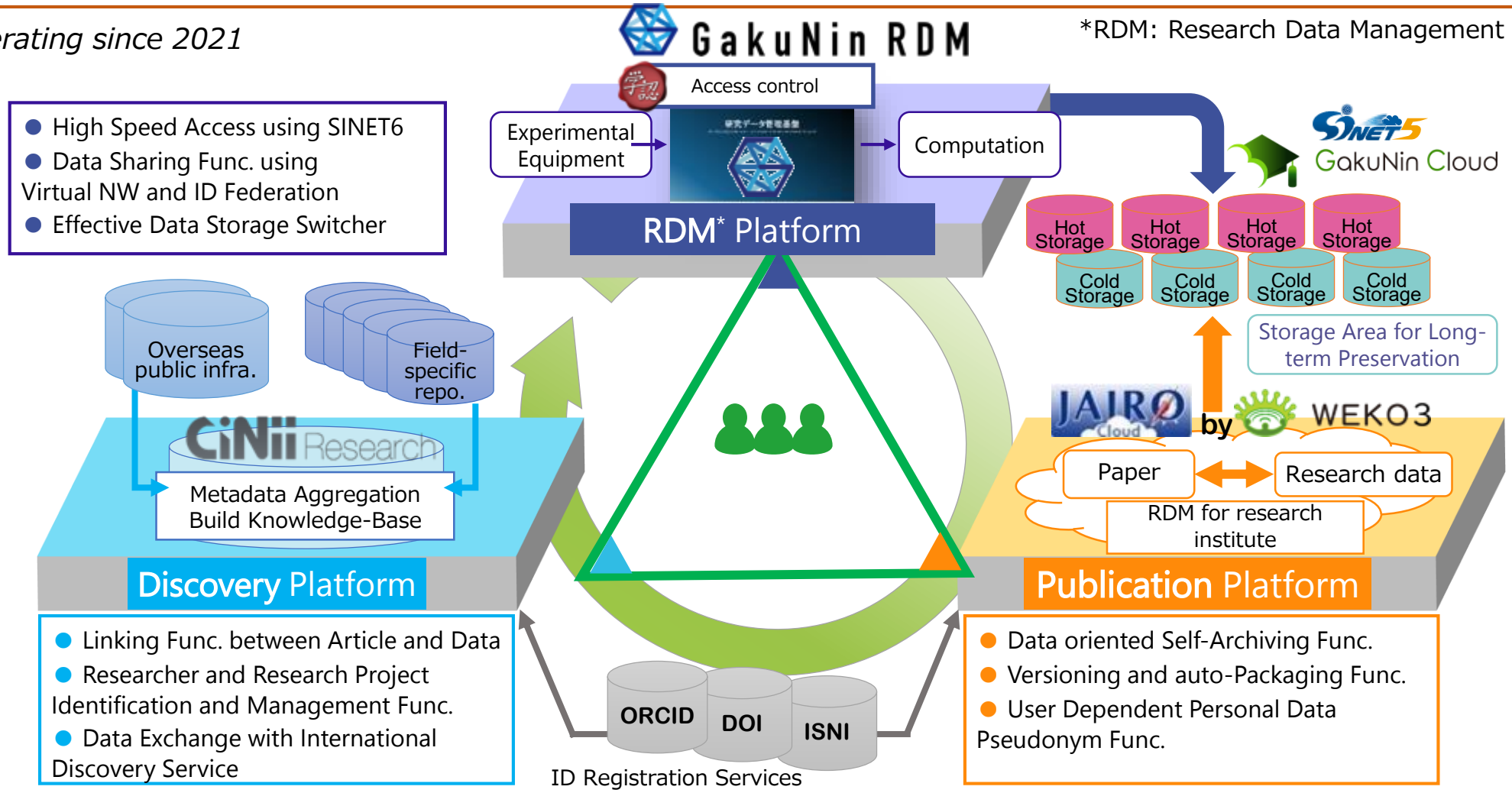
Research Divisions:
Seek to advance integrated research and development activities in information-related fields, including networking, software, and content.

Service Divisions:
Promote the creation of a state-of-the-art academic-information infrastructure, and support researchers and institutions.

NII Research Data Cloud (NII RDC)

Operating since 2021

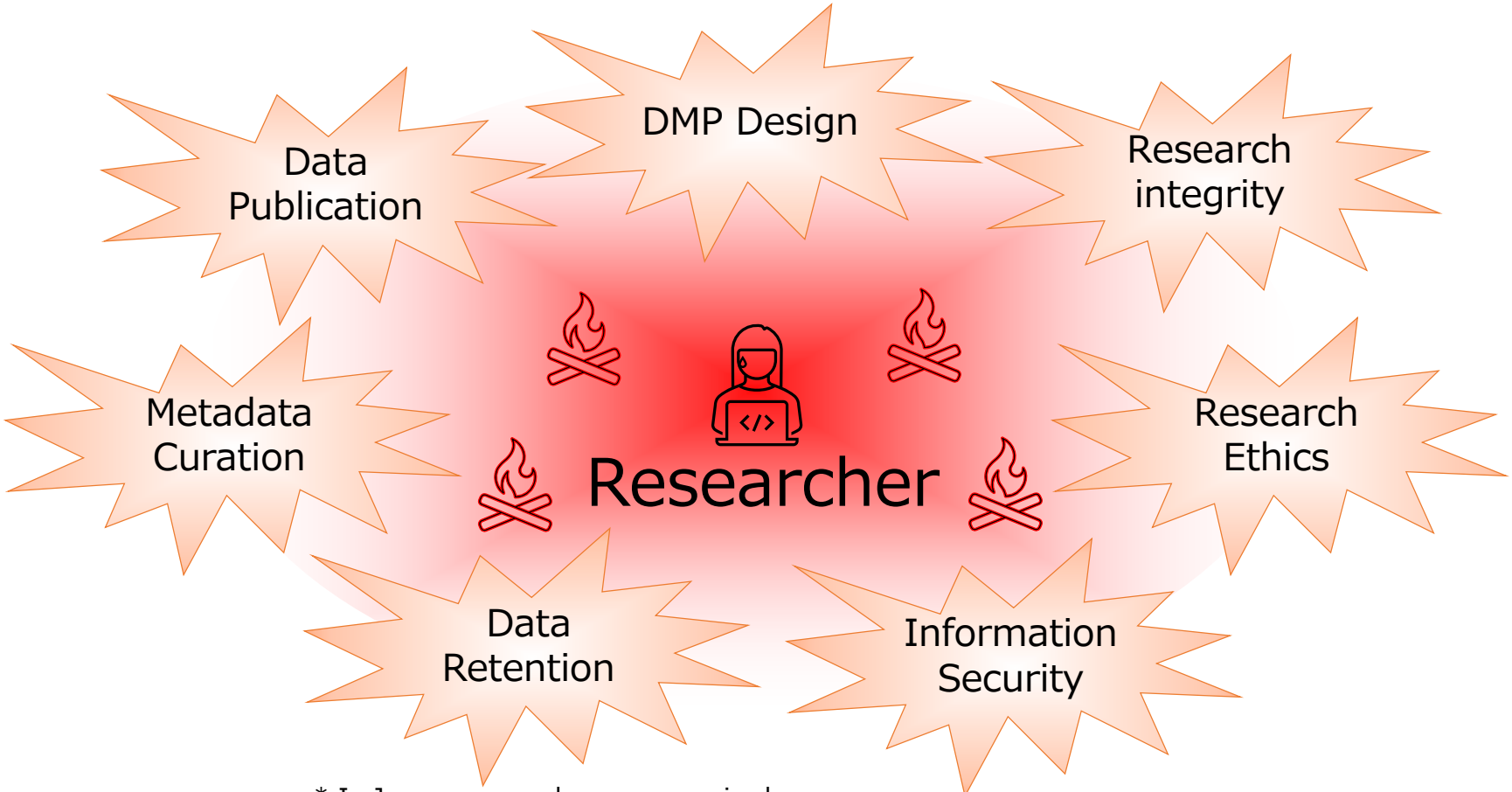
*RDM: Research Data Management



Mission: support researchers and institutions to carry out research activities and data management with research integrity and toward open science.

Increasing workload for researchers

So much extra work to do



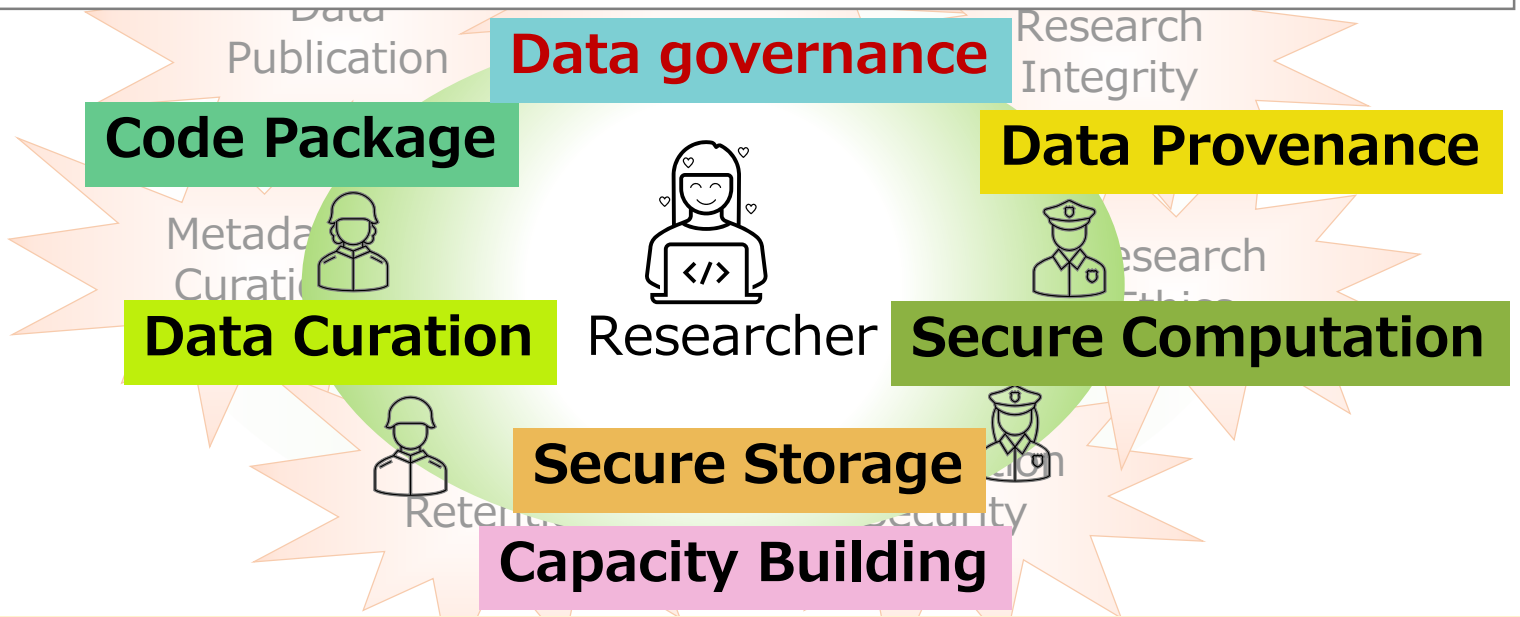
* In Japan, researchers are required to retain data for 10 years.

Seven new features of NII Research Data Cloud - shifting researchers' duties into motivation -

NII RCOS
Research Center for Open Science and Data Platform

Responsibilities of institution	Responsibilities of researcher
4-1 Establishing a data policy	6-1 Determining data to be managed
4-2 Appending metadata	6-2 Appending Metadata
4-3 Establishing human resources and support systems	6-3 Designing a DMP
4-4 Ensuring security	6-4 Retaining research data
	6-5 Publishing and sharing research data
	6-6 Post-project handling

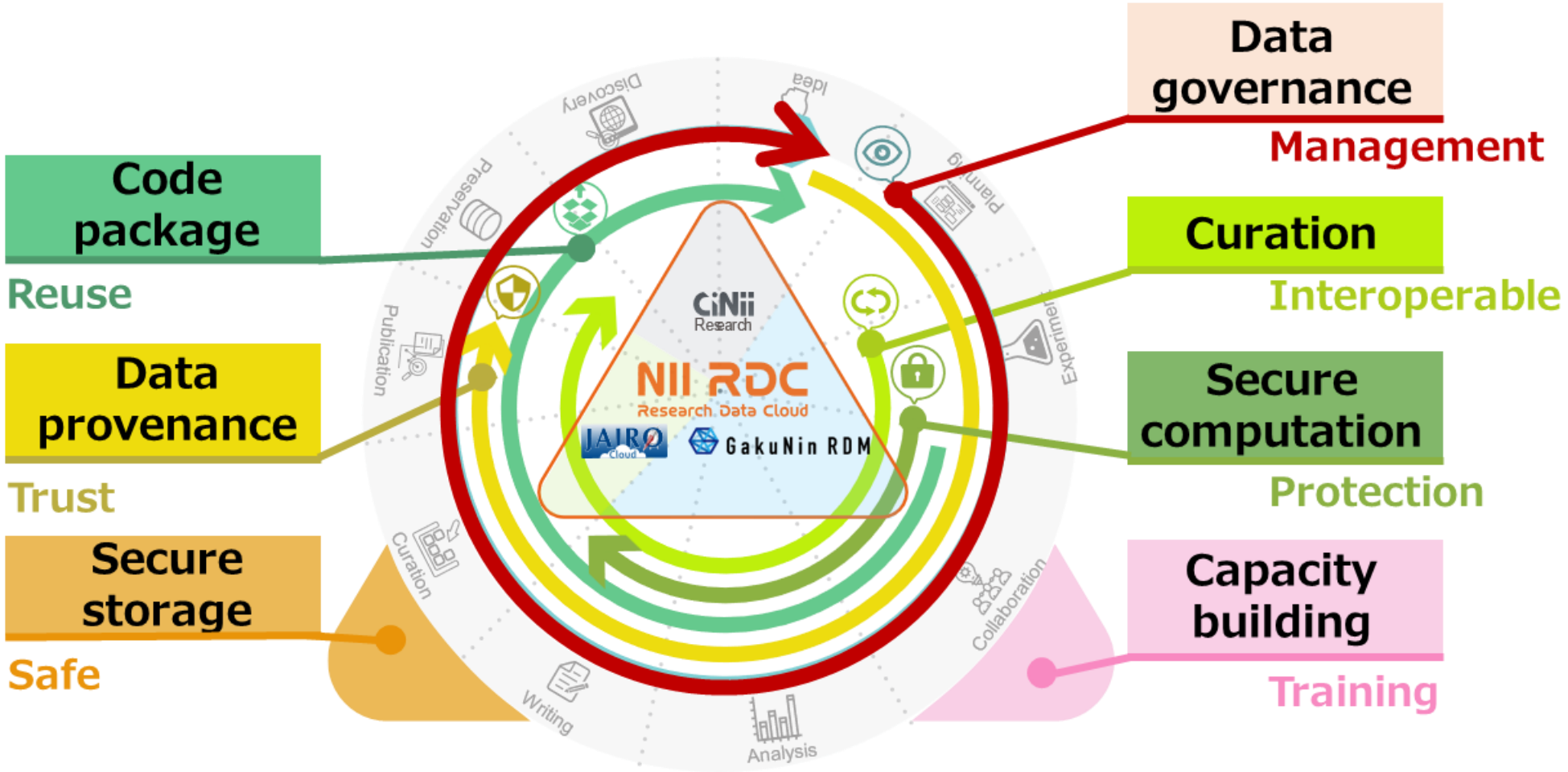
These items have been extracted from the Cabinet Office's "Basic Approach to the Management and Use of Publicly Funded Research Data".



Digital transformation is essential for researchers to turn 'extra work' into ‘a part of their research.'

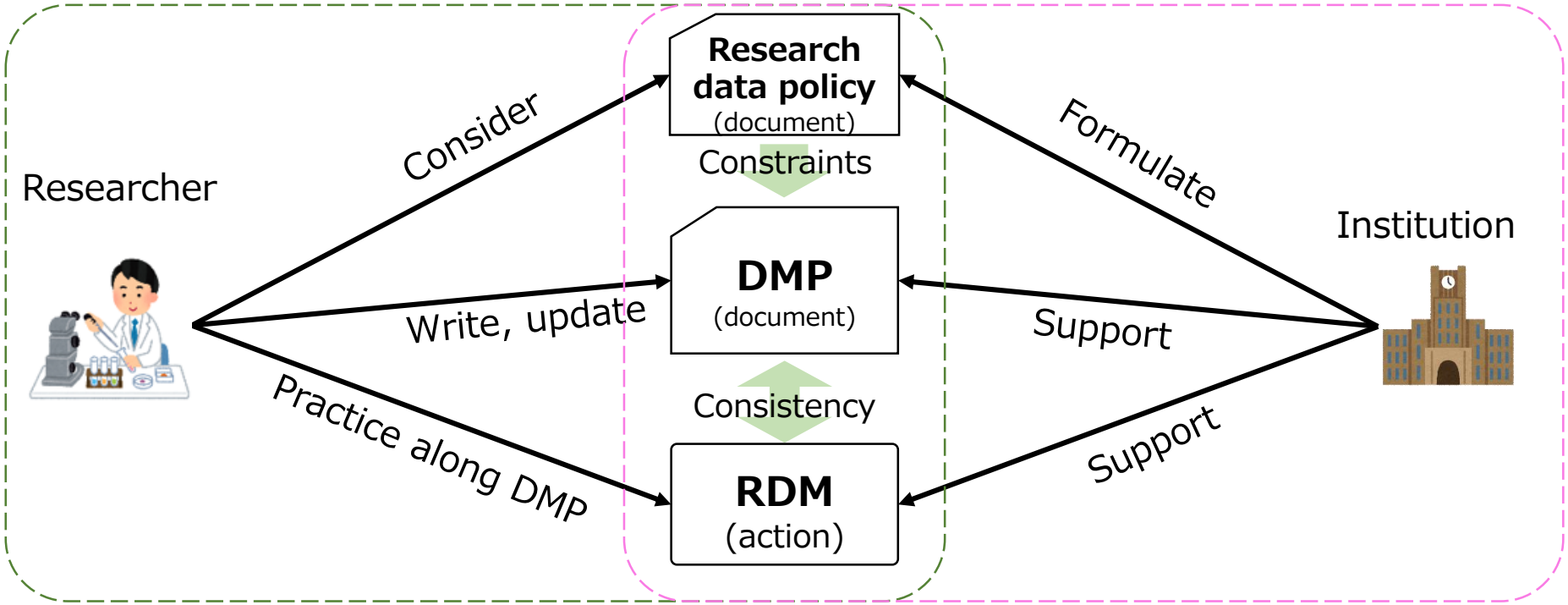
The role of data governance functions

Data governance (DG) functions support both researchers and institution staff to improve the quality of RDM.



Our research data governance (RDG) model

Definition of research data governance (RDG) in this talk:
Disciplined behavior/actions to improve the quality of RDM.



As a researcher:
- Consider the research data policy (RDP).
- Practice RDM along her/his DMP.
- Utilize the environment for RDM.
- Check the state of RDM along the DMP

As an institution:
- Formulate and maintain its RDP.
- Practice DG along its RDP.
- Support its researchers to practice RDM.
- Check the state of DMP along its RDP.

Challenges in support of DG as a researcher

The first purpose of our DG functions is to support researchers to carry out RDM along their requirements described in their DMPs.
We focus on three challenging points in terms of automatic support.

Researcher

Consider

Write, update

Practice along DMP

Research data policy (document)

Constraints

DMP (document)

Consistency

RDM (action)

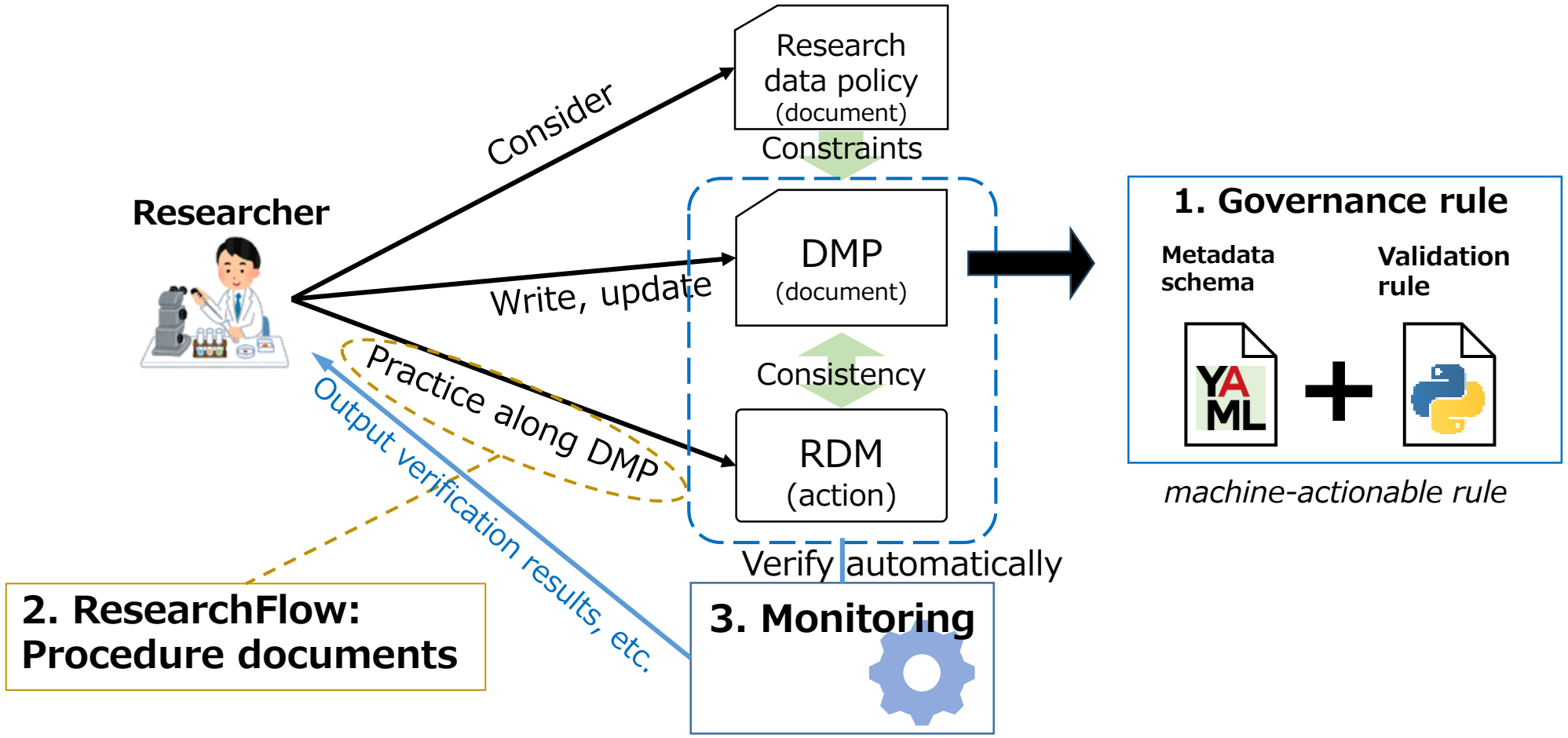
1. DMPs required by funding agencies are written basically **as free-form text**, which is difficult for a computation resource to interpret.

3. It's **burdensome** for researchers to verify whether the state of RDM satisfies the requirements in the DMP.

2. RDM according to the DMP is **left to researchers' discretion**.

Our first solution

DG as a researcher



Governance rule

Machine-actionable format of RDM requirements expressed by researchers, etc.

DMP
(document)

Conversion

In terms of verification:
- criteria on the state of RDM.

In terms of RDM procedures:
- requirements to the env (e.g., what they want to use).

Governance rule

Metadata schema:
Define schema of a (research data) entity

Validation rule(s):
Define a class including constraints on the metadata of the corresponding (research data) entity

Examples

```yaml
File:
  description: A file included in the research project, such as a text file, script file, or images.
  props:
    "@id":
      expected_type: str
      example: config/setting.txt
      required: Required.
      description: MUST be either a URI Path relative to the RO-Crate root directory or an absolute URI
    name:
      expected_type: str
      example: setting.txt
      required: Required.
      description: Denotes the file name.
    contentSize:
      expected_type: str
      example: 1560B
      required: Required.
      description: MUST be an integer representing the file size, suffixed with `B` for bytes. Other uni
```

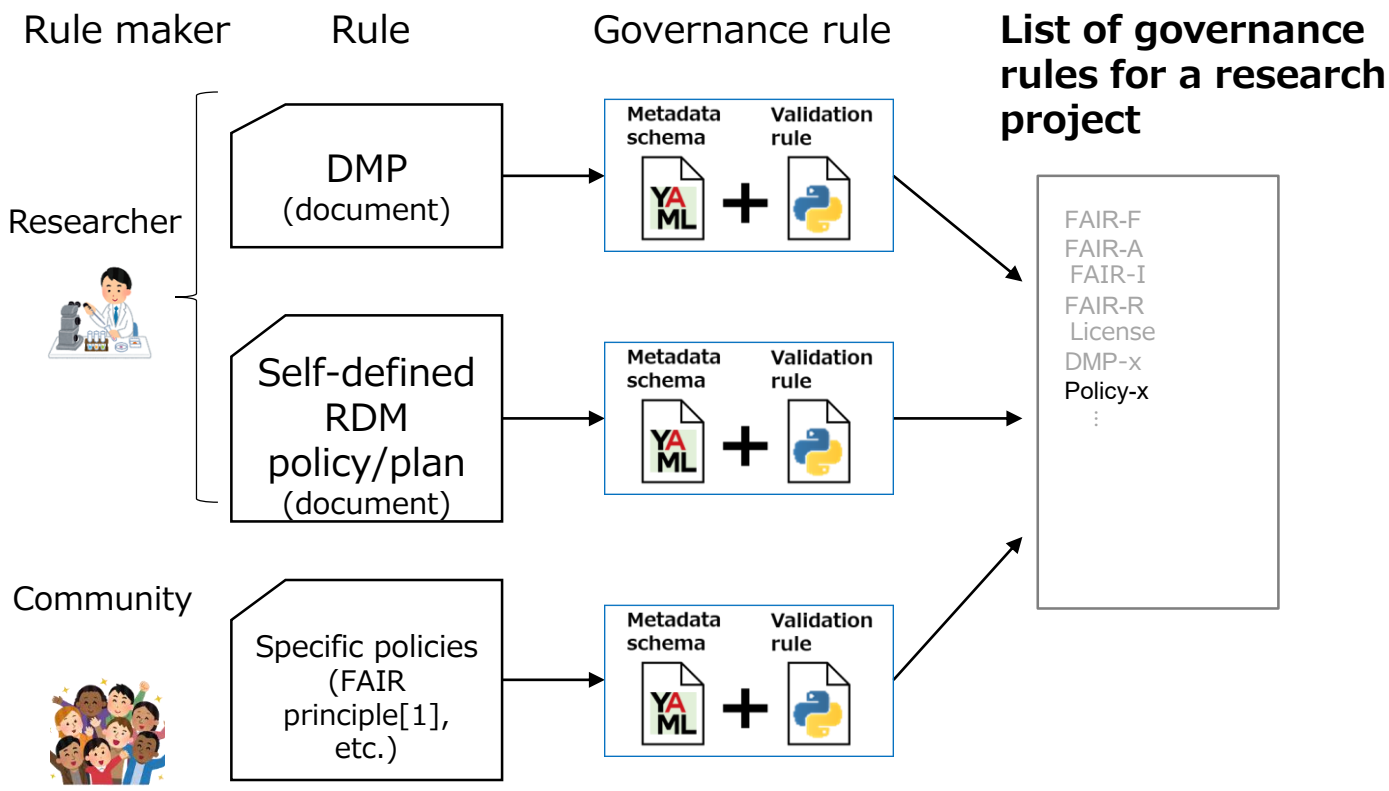
```python
class File(DataEntity):
    def __init__(
        self,
        id_: str,
        props: Dict[str, Any] = {},
        schema_name: str = SCHEMA_NAME,
        entity_def: EntityDef = SCHEMA_DEF["File"],
    ):
        super().__init__(id_, props, schema_name, entity_def)
```

```python
    def validate(self, crate: "ROCrate") -> None:
        super().validate(crate)

        error = EntityError(self)

        if is_url(self.id):
            if "sdDataPublished" not in self:
                error.add(
                    "sdDataPublished",
                    "The property `sdDataPublished` is required when the id is a URL.",
                )
```

Quoted from https://github.com/NII-DG/nii-dg/nii_dg/schema/base.yml and base.py

Extension of governance rules

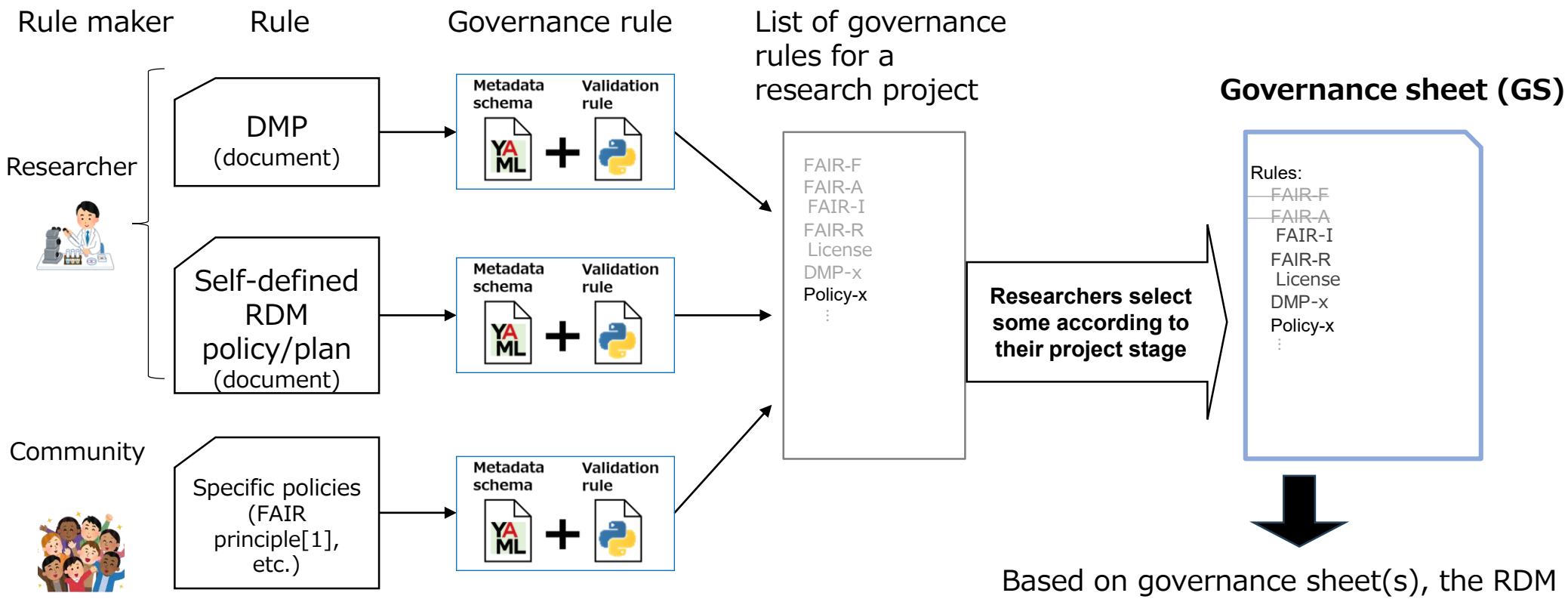
Extensible so that researchers can set requirements specific to each project.



[1] https://force11.org/info/the-fair-data-principles/

Governance sheet

A list of governance rules selected for a research project



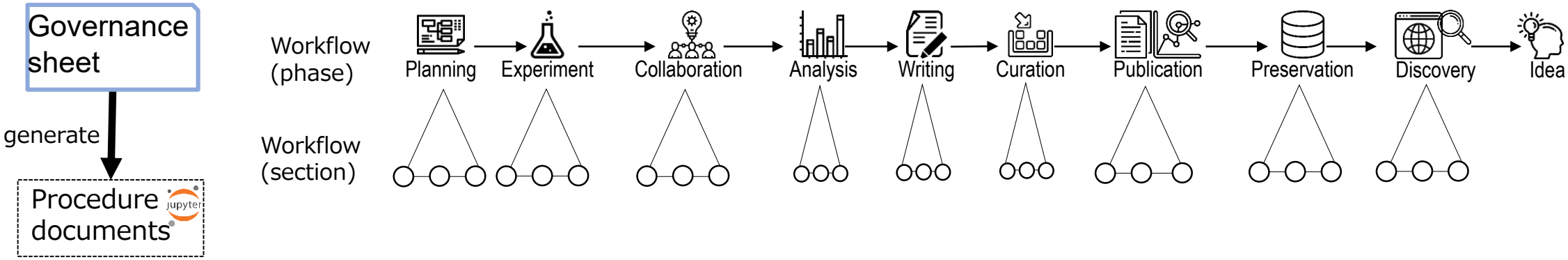
Based on governance sheet(s), the RDM system

- generates procedure documents for RDM (ResearchFlows),
- verifies the state of RDM (monitoring).

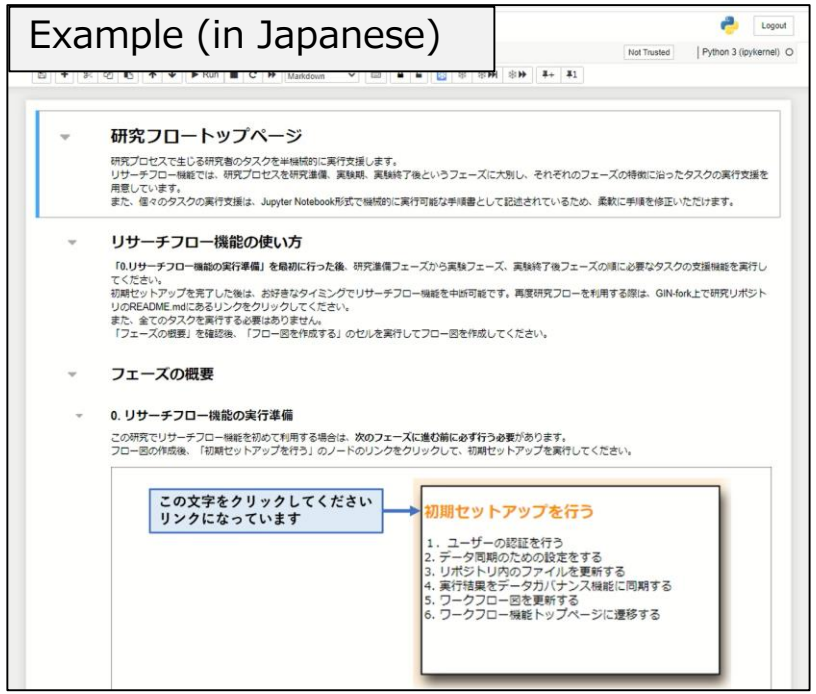
[1] https://force11.org/info/the-fair-data-principles/

ResearchFlow

Actionable procedure documents based on a governance sheet



Example (in Japanese)

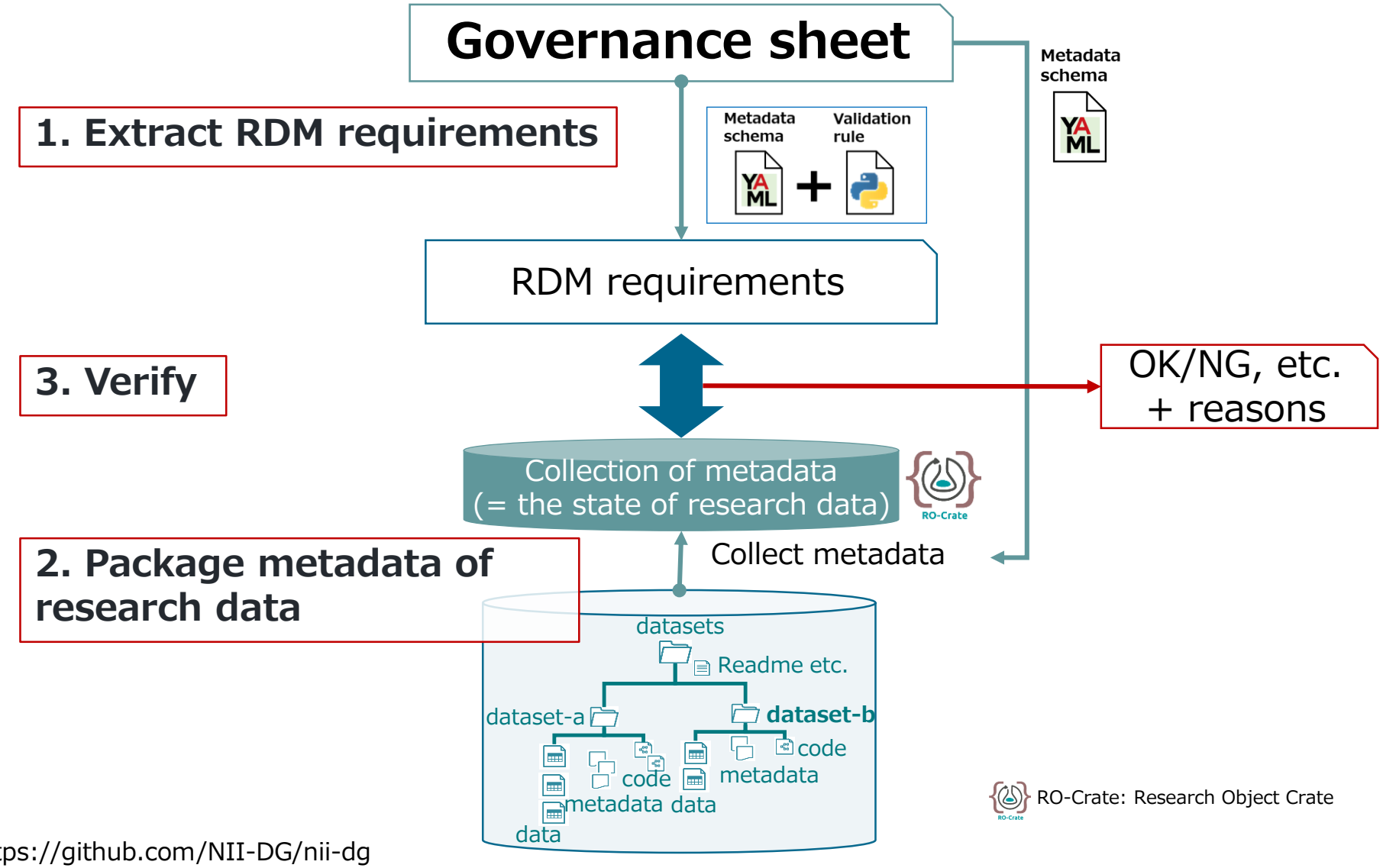


ResearchFlows can be

- modified/extended by researchers so that they pave “their own RDM way”.
- used like electronic laboratory notebooks.

Monitoring

Verify whether the RDM state satisfies the requirements described in the Governance sheet

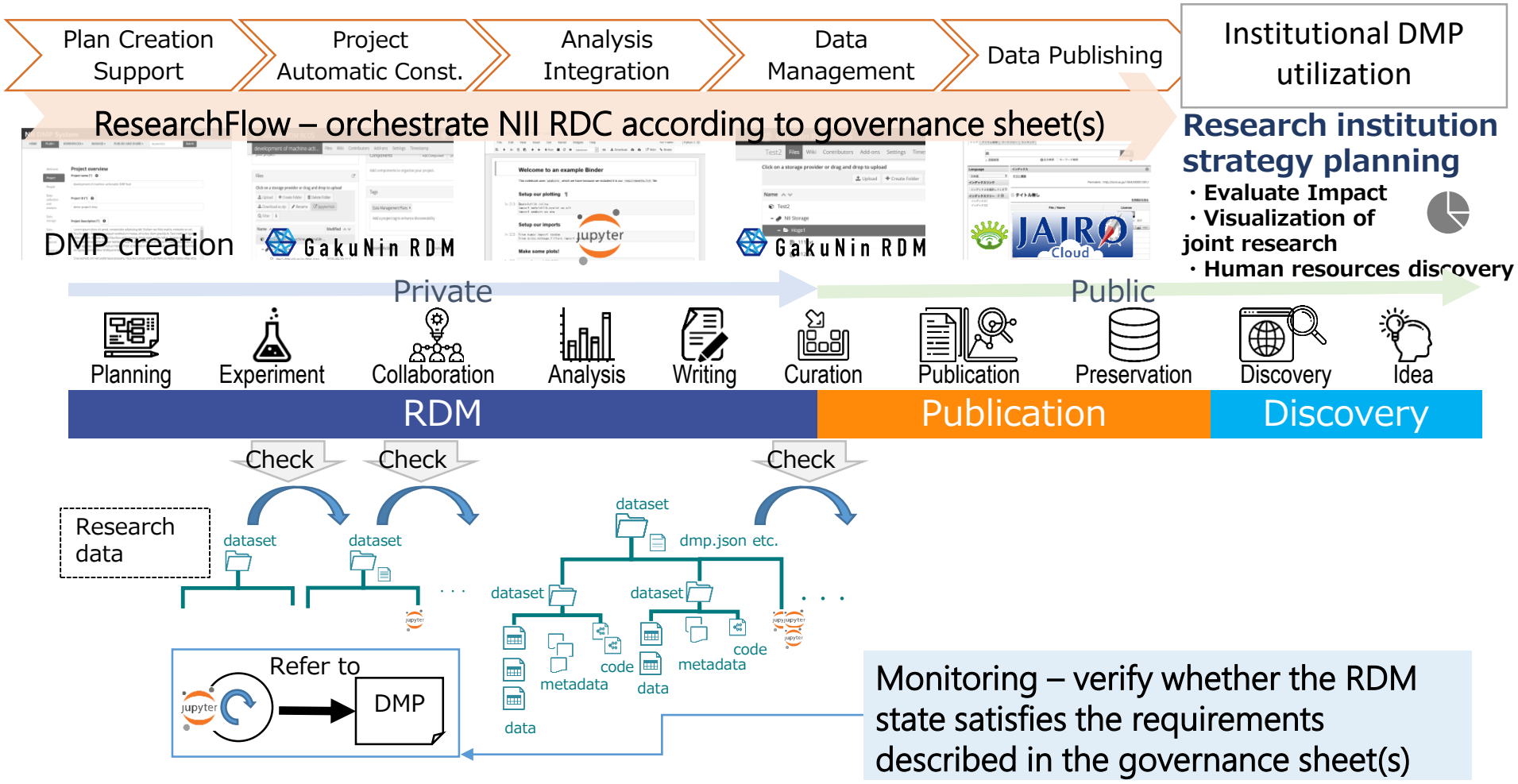


RO-Crate: Research Object Crate

https://github.com/NII-DG/nii-dg

Our vision of the data governance functions

Automation of data management process based on governance sheets



Summary

- Research data management (RDM) has become more and more important in terms of the emphasis on scientific integrity and open science.
- We have been developing data governance functions to support researchers in carrying out RDM in their own way.
  - Governance rule & sheet: express RDM requirements in a machine-actionable format.
  - ResearchFlow: arrange the RDM procedure and a suitable RDM environment based on governance sheet(s).
  - Monitoring: verify whether the state of RDM satisfies the RDM requirements described in governance sheet(s).