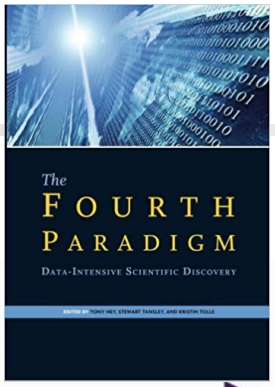
NII RCOS
Research Center for Open Science and Data Platform

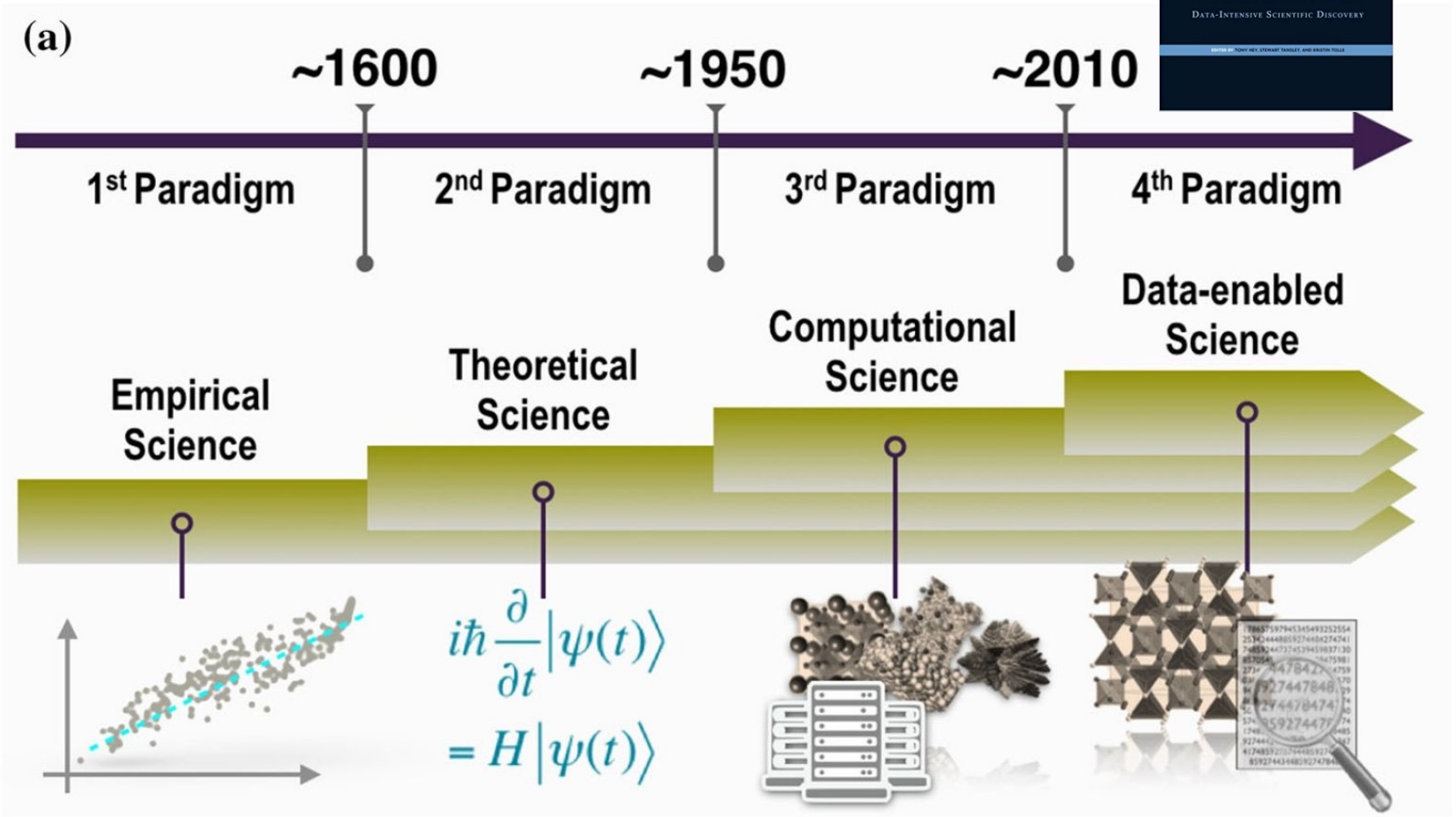
Trends in Research Data Management

Kazu YAMAJI
National Institute of Informatics

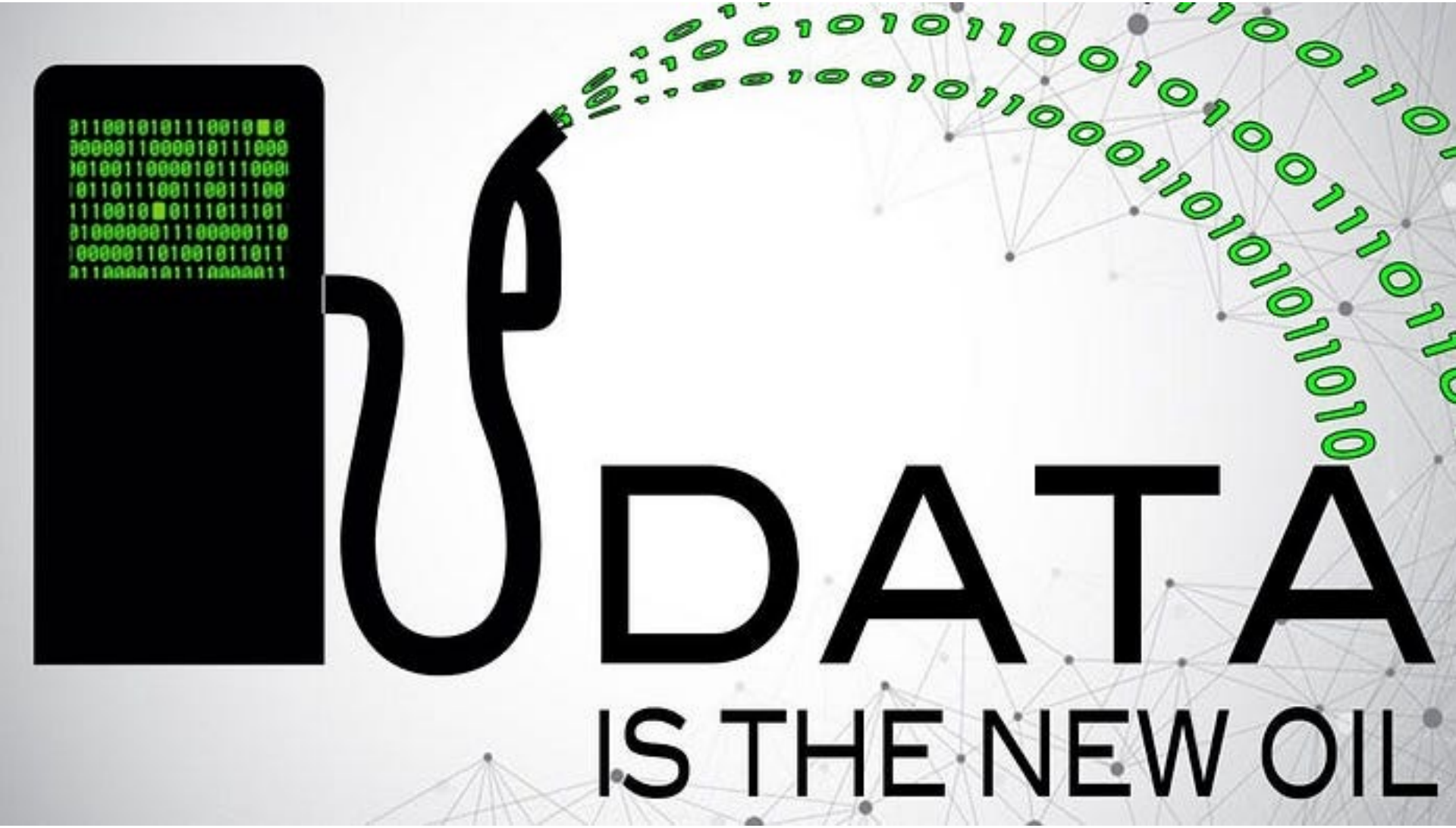
LIBSENSE Workshop 2024
4th March 2024

4th Paradigm



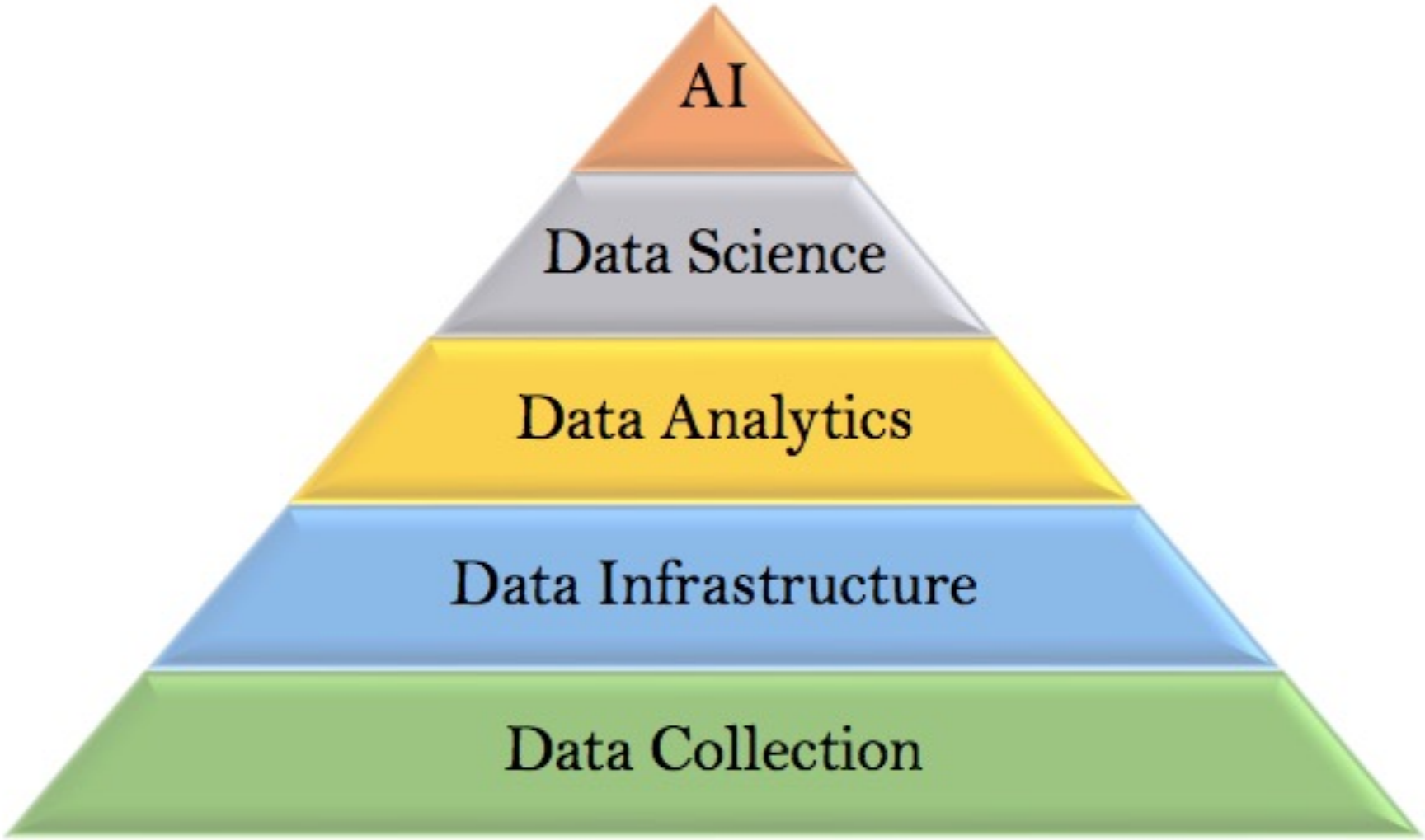


Pilania, G., Liu, XY. & Wang, Z. J Mater Sci (2019) 54: 8361. https://doi.org/10.1007/s10853-019-03434-7



https://medium.com/@adeolaadesina/data-is-the-new-oil-2947ed8804f6

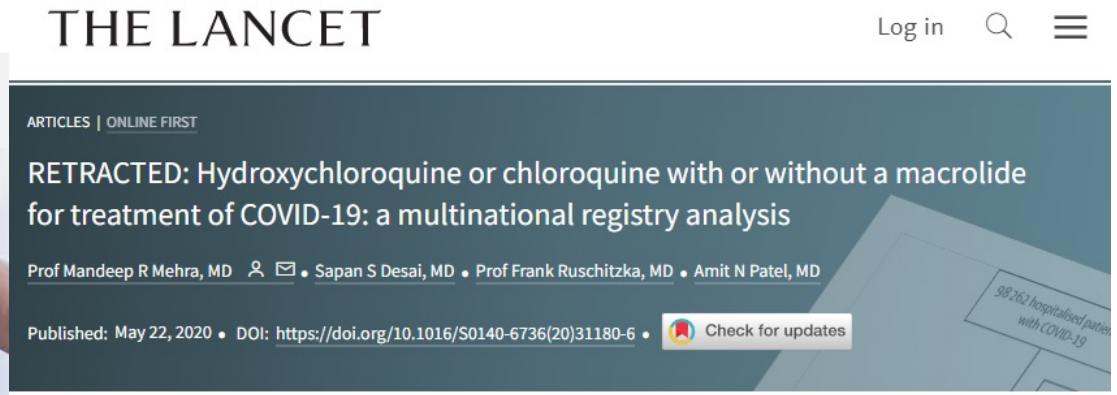
The AI Hierarchy of Needs

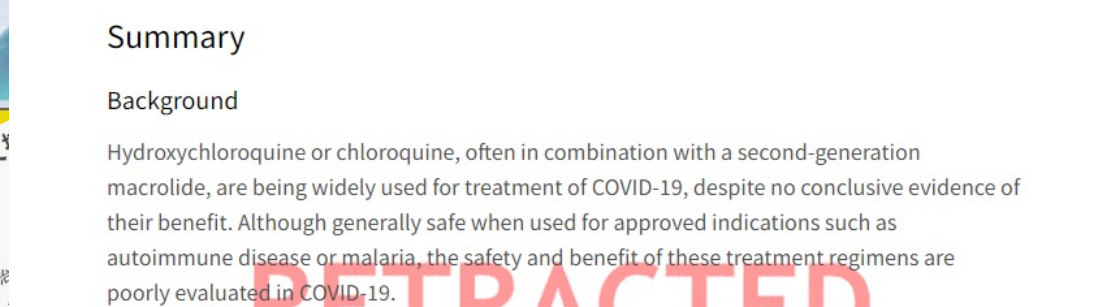


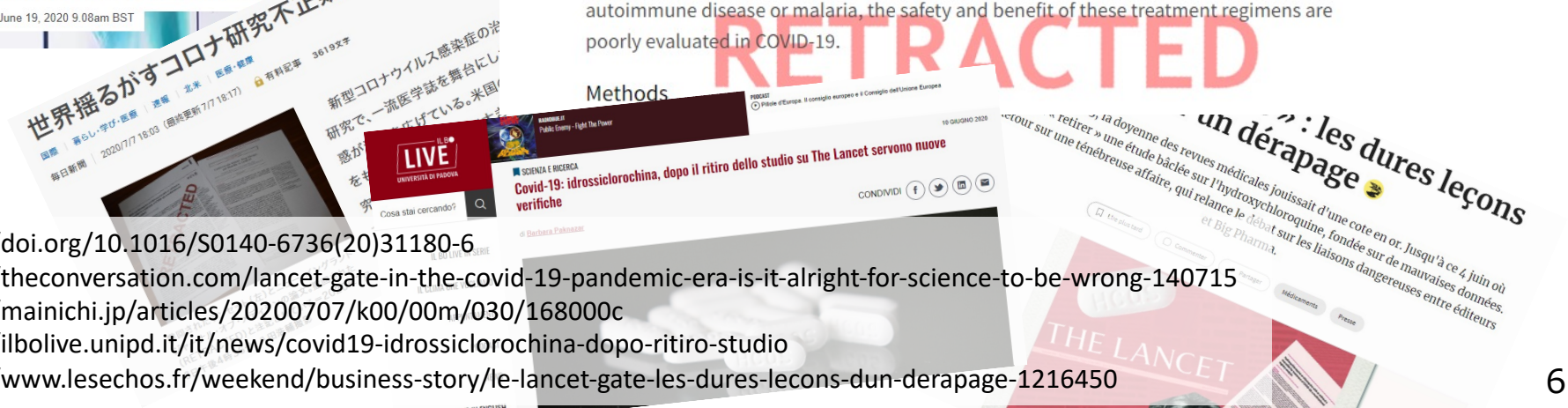
https://medium.com/swlh/the-ai-hierarchy-of-needs-9b015d061f29

https://www.theom.ai/blog-posts/ai-needs-data-and-data-needs-governance

Lance gate in the COVID-19 Pandemic



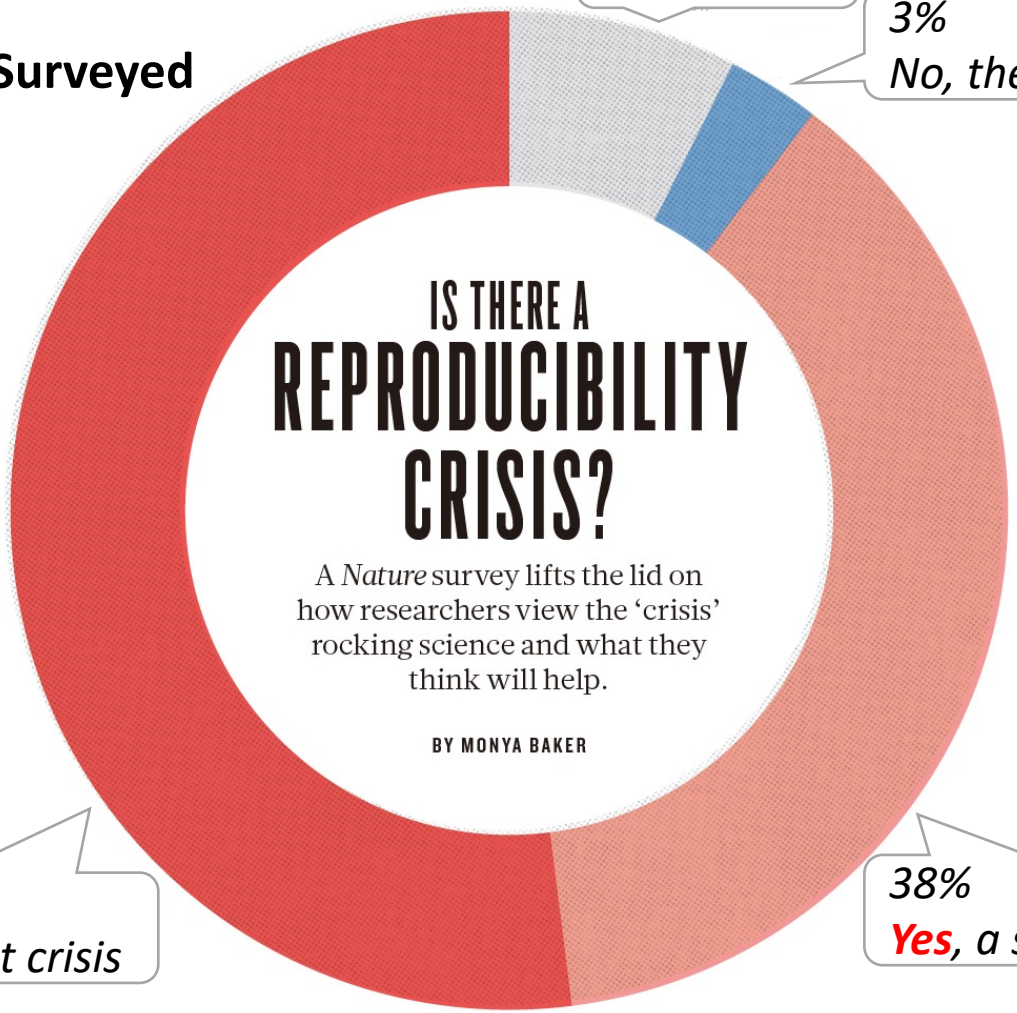




https://doi.org/10.1016/S0140-6736(20)31180-6
https://theconversation.com/lancet-gate-in-the-covid-19-pandemic-era-is-it-alright-for-science-to-be-wrong-140715
https://mainichi.jp/articles/20200707/k00/00m/030/168000c
https://ilbolive.unipd.it/it/news/covid19-idrossiclorochina-dopo-ritiro-studio
https://www.lesechos.fr/weekend/business-story/le-lancet-gate-les-dures-lecons-dun-derapage-1216450

Reproducibility Crisis

1,576
Researchers Surveyed



7%
Don't know

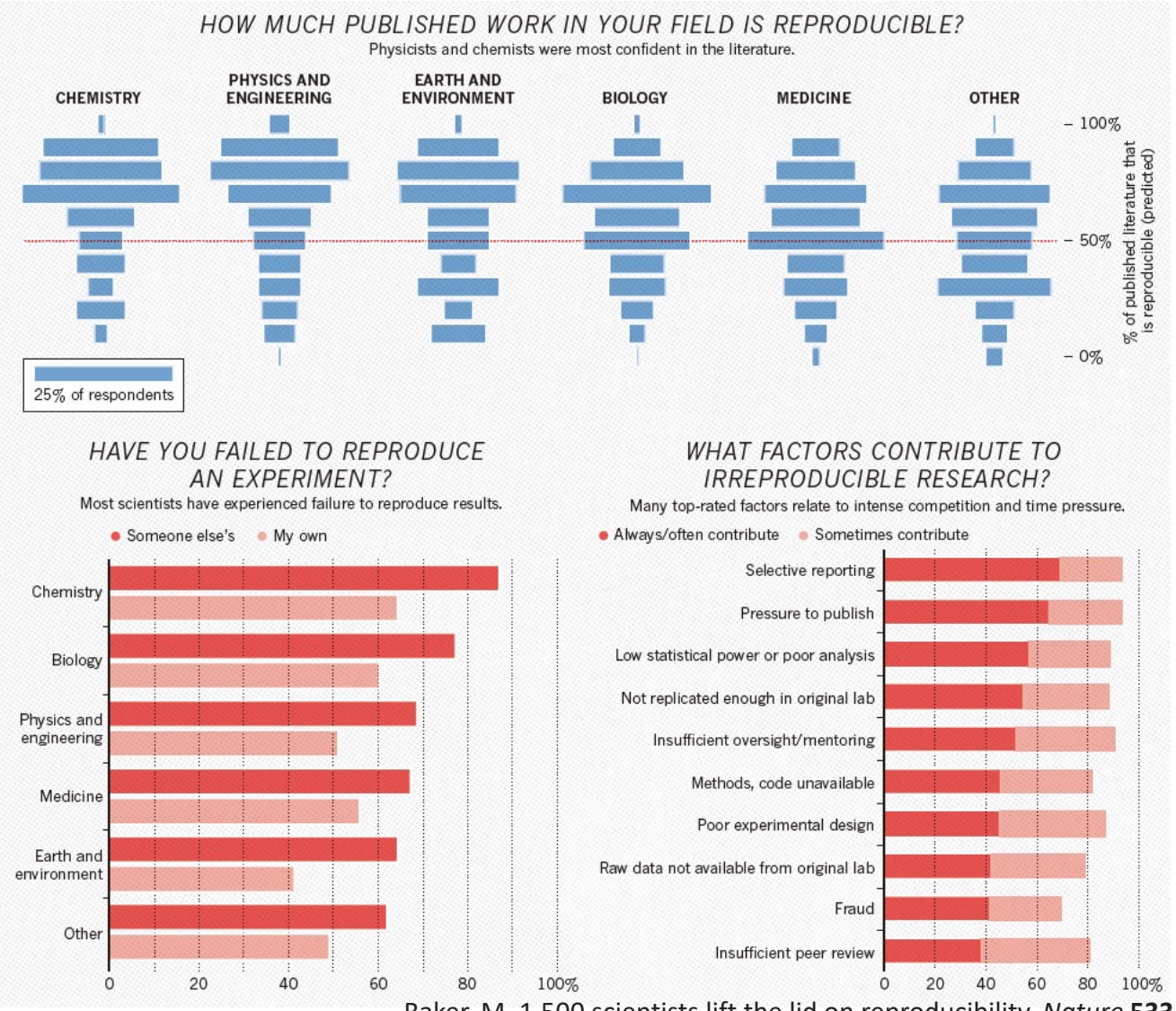
3%
No, there is no crisis

52%
***Yes**, a significant crisis*

38%
***Yes**, a slight crisis*

Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533**, 452–454 (2016). https://doi.org/10.1038/533452a

Reproducibility Crisis



Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533**, 452–454 (2016). https://doi.org/10.1038/533452a

Recent Trends in Research Data Management

18th International Digital Curation Conference

19-21 February 2024 |

IDCC24 Silver Sponsor Artefactual Systems Inc.

We are excited to announce Artefactual Systems Inc. are sponsoring the 18th International Digital Curation Centre.

You can learn more about their work via their website.

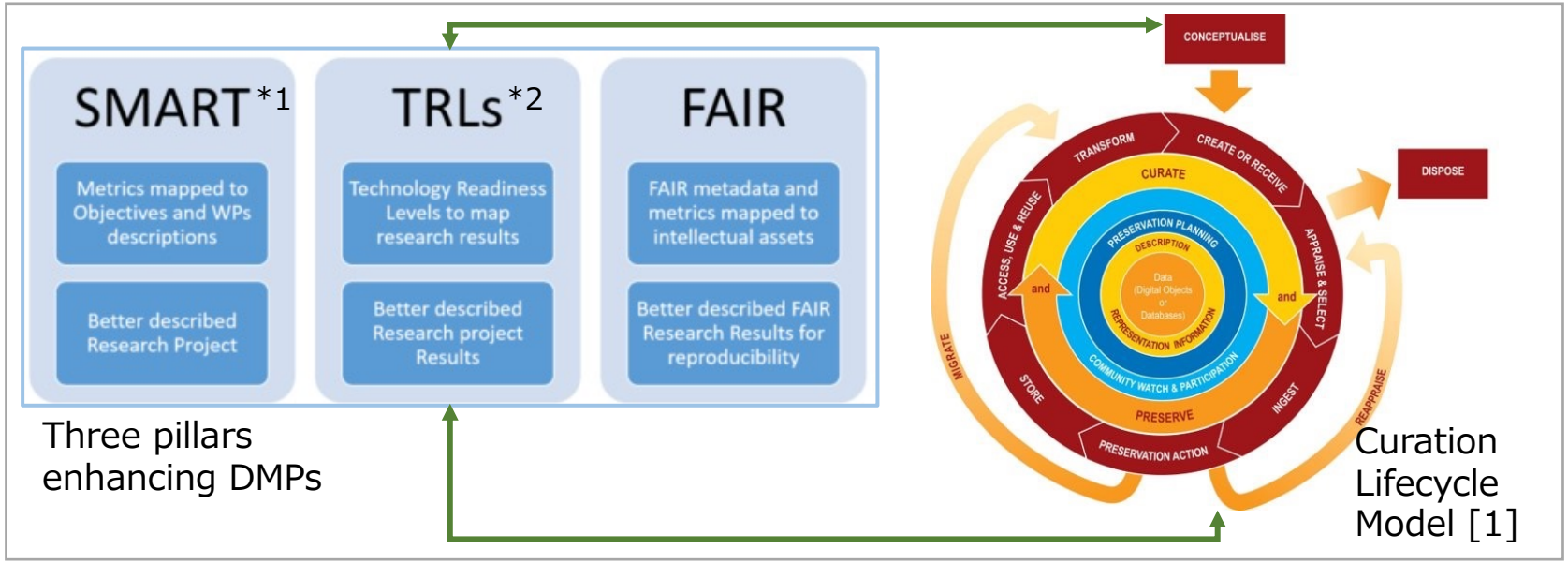
Accepted Submissions
Accommodation
Call for Papers
Frequently Asked Questions
Key Dates
Programme
Registration
Social Events

https://www.dcc.ac.uk/events/idcc24

Trend 1: "DMPs as Management Tool for Intellectual Assets by SMART-metrics" [1]

(in Univ. Of Bern)

As Is: Just funders' requirements
To Be: ***a valuable tool for research project management that helps researchers optimize the planning, documenting, and executing of Open science***



- Enhanced DMPs enable researchers to manage research data as their **intellectual assets**.
- Curation efforts should be during research data life cycle and the whole research process.

*1: Specific, Measurable, Attainable, Relevant, Time-based
*2: Technology Readiness Levels

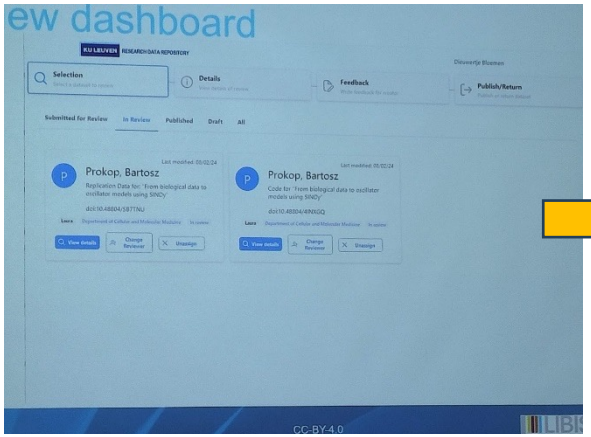
[1] F. G. Toro, presentation in 18th IDCC (20th Feb. 2024)
[2] https://www.dcc.ac.uk/guidance/curation-lifecycle-model

Trend 2: Developing a dashboard for reviewing datasets to comply with funders' requirements

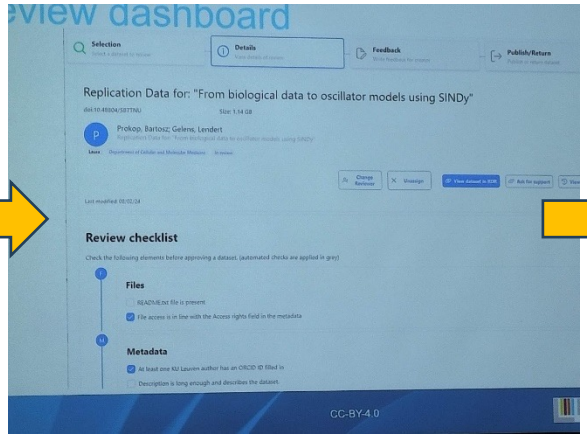
Extend KU Leuven's institutional data repository to create the review dashboard for helping researchers to teach FAIR and data sharing best practices

Key features:

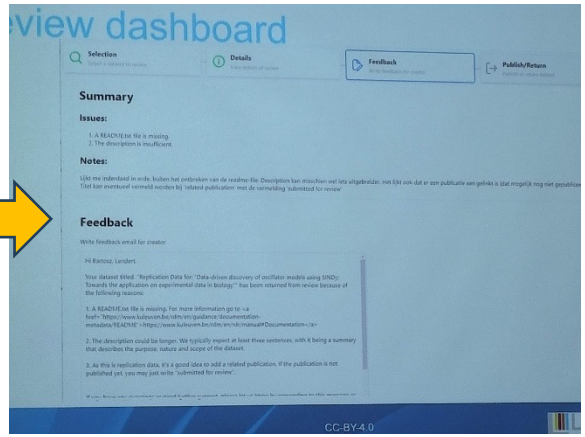
① Assignment of reviewers & auto-assignment for returning datasets
② Checklist of items to review & notes field
③ Autogenerated feedback based on checklist
④ Storage tracking
⑤ Logs to retain the history of a dataset's review phases
⑥ Uniform messaging & easy follow up of communication



Select data for review



Review based on checklist



Auto-generate feedback comment

Open Science E-Infra in JAPAN

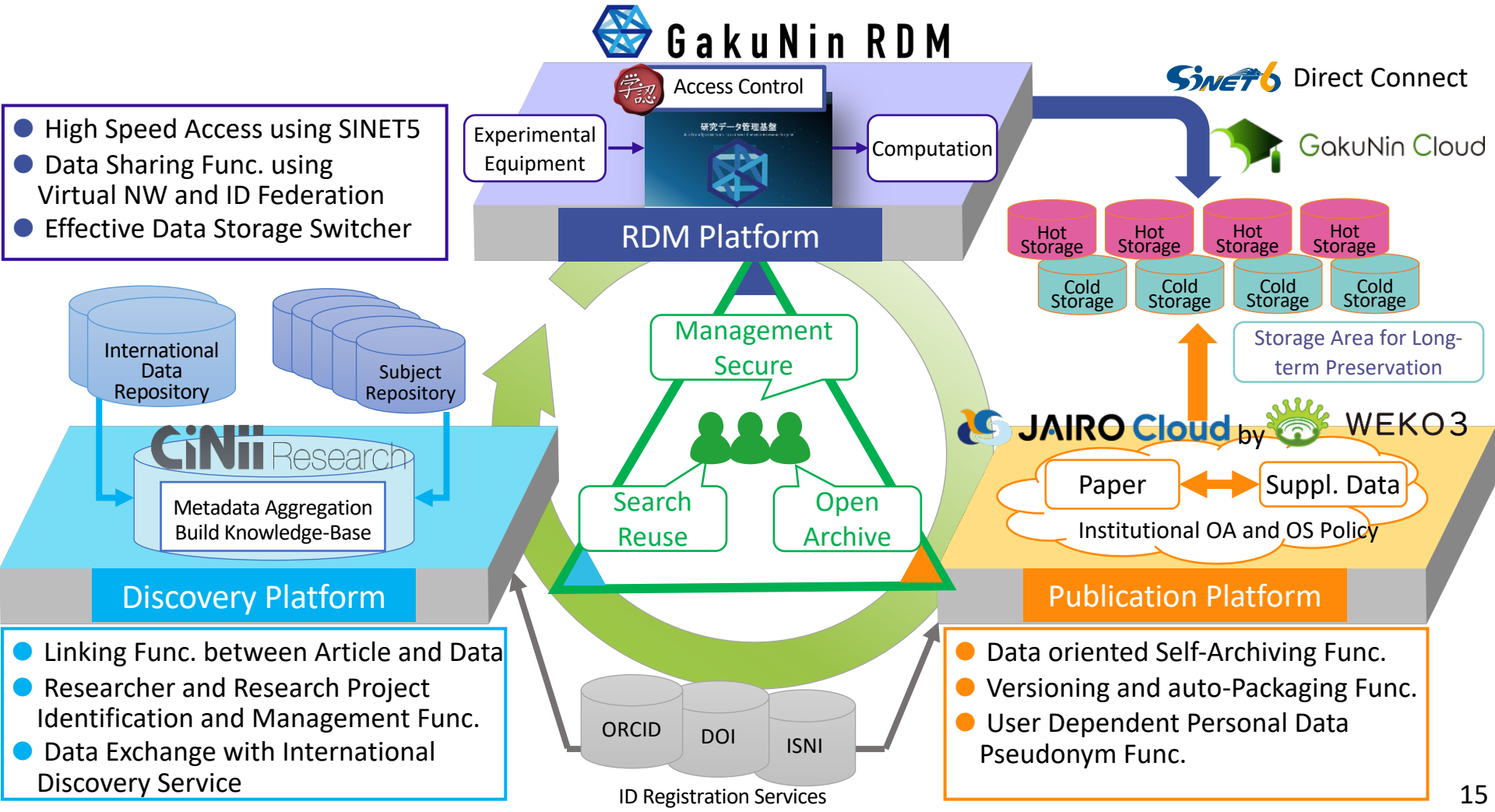
NII Research Data Cloud

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)

NII Research Data Cloud

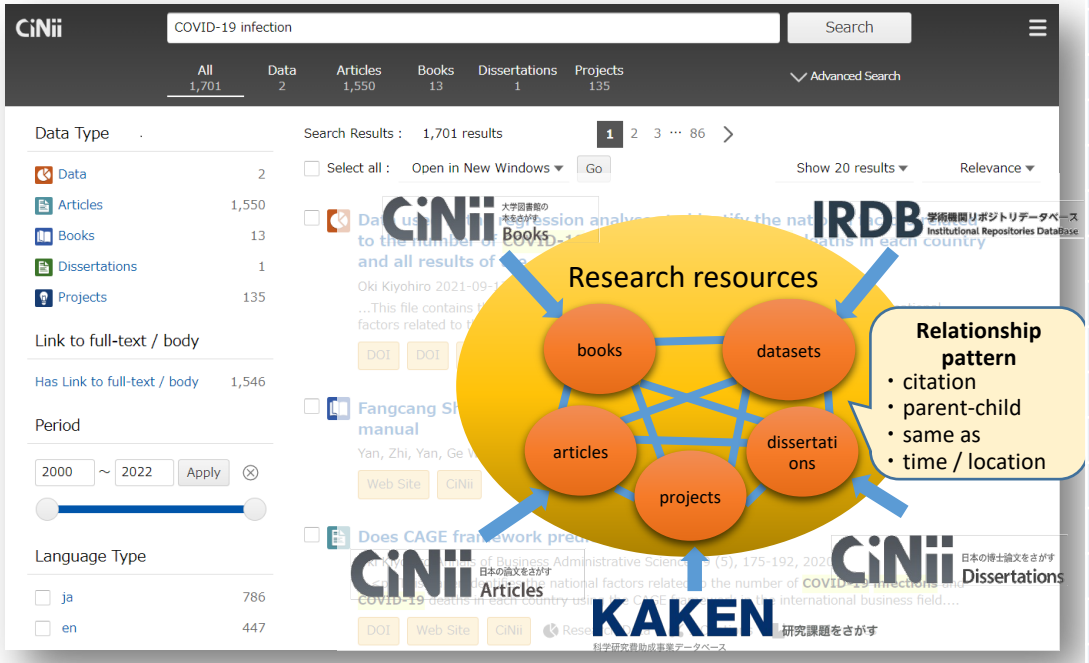
2017 Start Development ⇒ **2021** Start Operation



CiNii Research: Discovery Platform

- Provide search engines for **various research outputs**.
  (research datasets, articles, books, dissertations, projects, etc.)
- Develop a huge academic **knowledge graph**.

CiNii Research



Number of Entities

Data Sources	Research Outputs	Researchers	Projects
CROSSREF	10,344,535	63,705,306	0
JALC	11,137,631	18,783,938	0
KAKEN	2,160,330	513,620	1,032,388
IRDB	1,784,754	2,619,821	0
CIB	13,011,327	8,299,962	0
CID	714,435	701,885	0
NDL	12,350,521	2,172,722	0
NINJAL	391	948	0
IDR	961	3,411	0
RUDA	70	75	0
ERAD	0	177,128	0
INTEGBIO	1,917	817	0
JDCAT	5,828	23,258	0
PUBLIC_DATA	104,153	0	0
LSDB Archive	852	497	0
MDR	14,295	3,751	0
PubMed	35	654,599	0

Collecting more sources

JAIRO Cloud: Publication Platform

- Expand functions to accelerate to promote **Open Access**
- Provide functions to support **Research Data Publication**

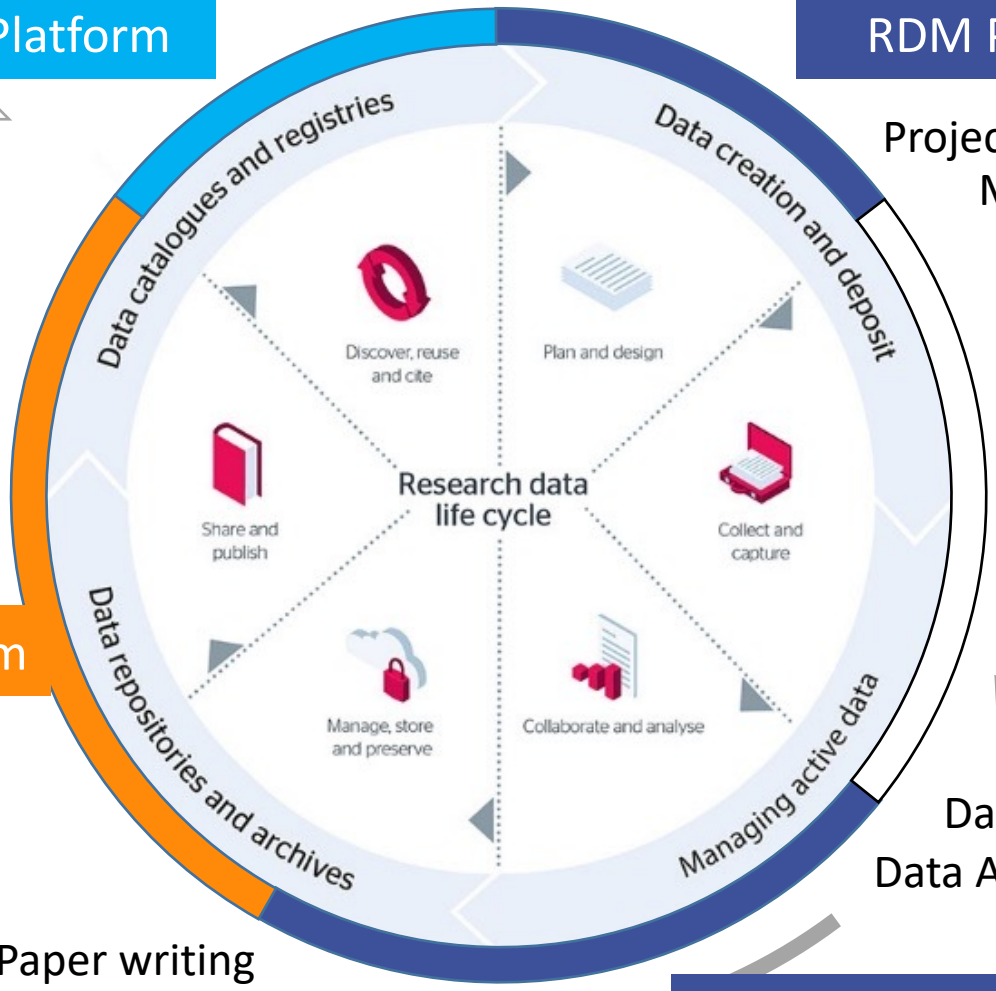
Migration to new system is completed!

Use of the Social Sciences and Humanities field

Use in other research fields

- Life Science (RIKEN)
- Medicine (AMED)
- Meteorological Science (Polar Research Institute)
- Multidisciplinary Science (mdx: Data Platform project, University of Tokyo)
- etc.

Relationship between Research Data Infrastructure and Research Workflow



Discovery Platform

RDM Platform

Project Start (Application)
Member Management
Initial Setting

Aggregation

Experiment
Data Acquisition

Publication Platform

Institutional or
Domain
Repository

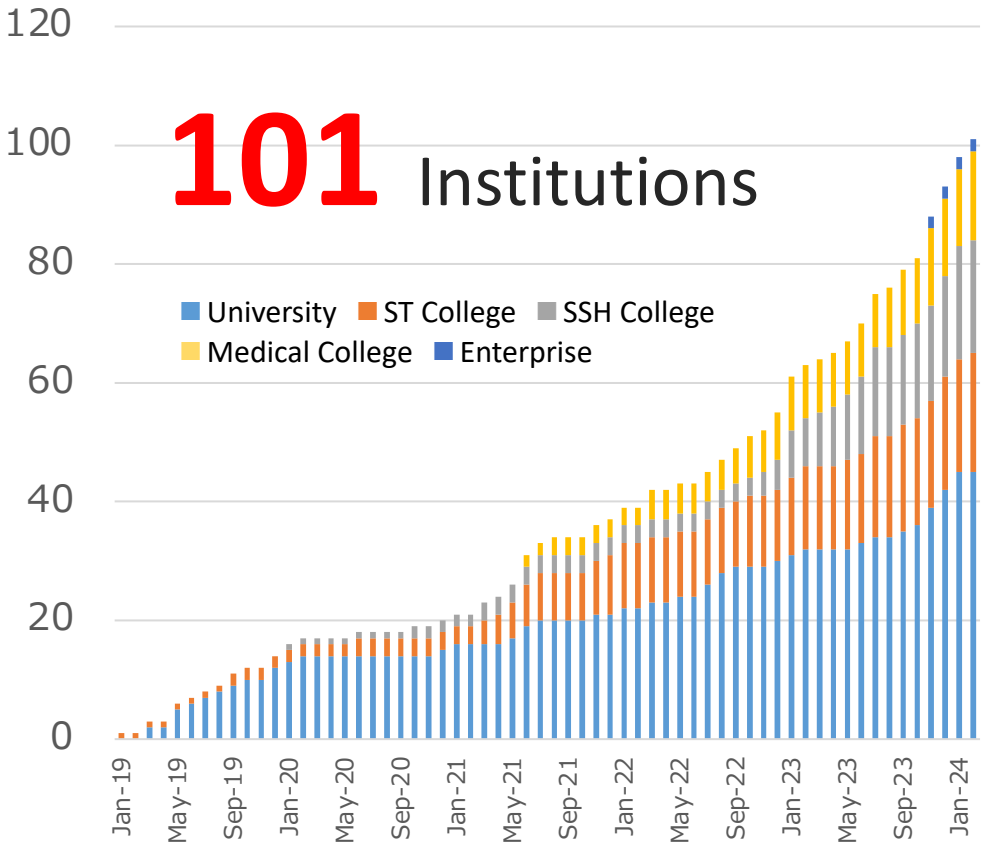
Data Management
Data Analysis

Paper writing
Deposit with Supplemental Data

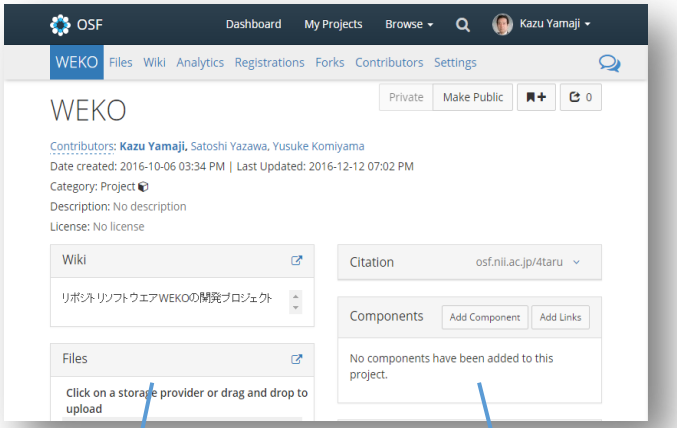
RDM Platform

GakuNin RDM: Research Data Management Platform

- Provide the RDM portal function that is easy for **researchers** to use.
- Provide **organizational** RDM with their data policy.



GakuNin RDM



Research Software

Institutional Storage

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.

Management
Data Governance

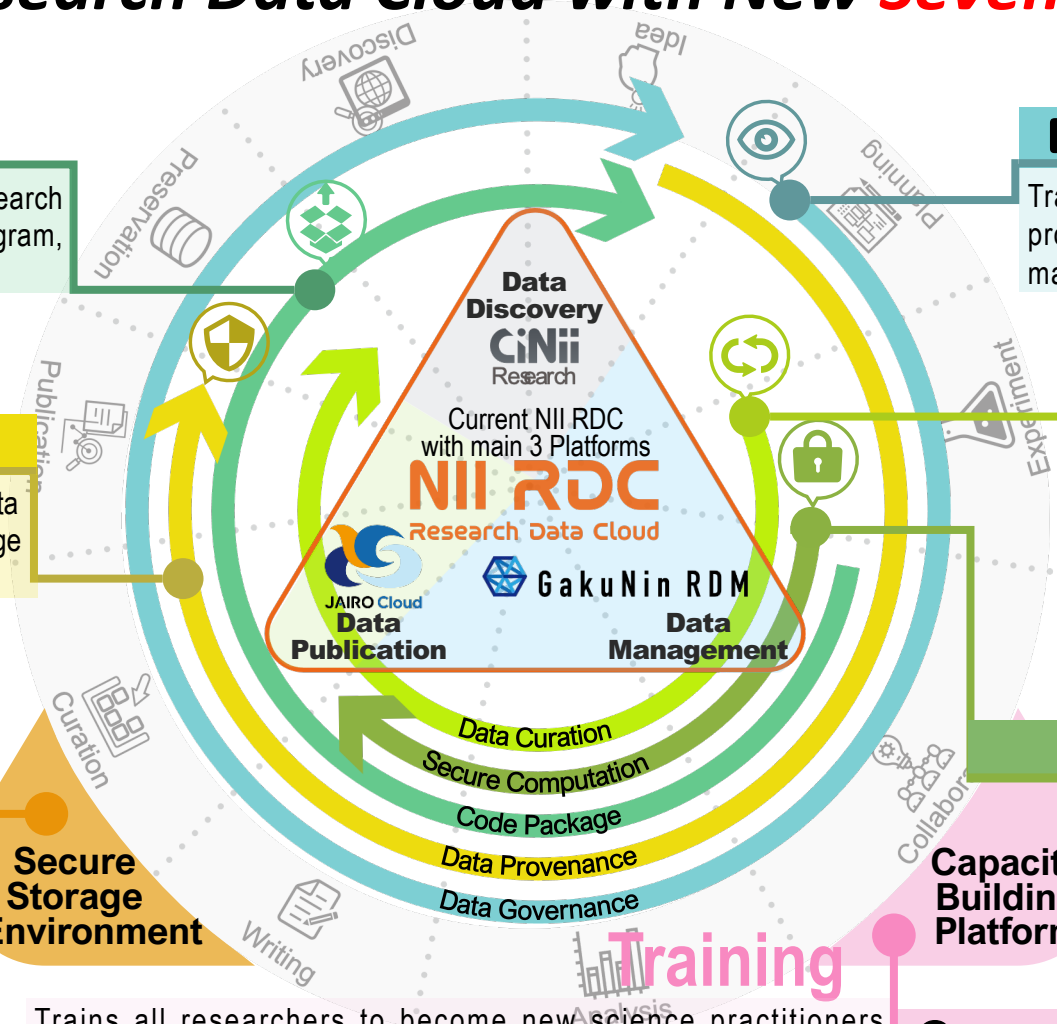
Transforms DMP into a crucial project document by supporting machine-actionable functions.

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.



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

Capacity Building

Data Management Plan

DCC

Checklist for a Data Management Plan, v4.0

Please cite as: DCC. (2013). *Checklist for a Data Management Plan.* v.4.0. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/data-management-plans

DCC Checklist	DCC Guidance and questions to consider
Administrative Data	
ID	A pertinent ID as determined by the funder and/or institution.
Funder	State research funder if relevant
Grant Reference Number	Enter grant reference number if applicable [POST-AWARD DMPs ONLY]
Project Name	If applying for funding, state the name exactly as in the grant proposal.
Project Description	**Questions to consider:** - What is the nature of your research project? - What research questions are you addressing? - For what purpose are the data being collected or created? **Guidance:** Briefly summarise the type of study (or studies) to help others understand the purposes for which the data are being collected or created.
PI / Researcher	Name of Principal Investigator(s) or main researcher(s) on the project.
PI / Researcher ID	E.g ORCID http://orcid.org/
Project Data Contact	Name (if different to above), telephone and email contact details
Date of First Version	Date the first version of the DMP was completed
Date of Last Update	Date the DMP was last changed
Related Policies	**Questions to consider:** - Are there any existing procedures that you will base your approach on? - Does your department/group have data management guidelines? - Does your institution have a data protection or security policy that you will follow? - Does your institution have a Research Data Management (RDM) policy? - Does your funder have a Research Data Management policy? - Are there any formal standards that you will adopt? **Guidance:** List any other relevant funder, institutional, departmental or group policies on data

Vision: Data Governance Function

Issue

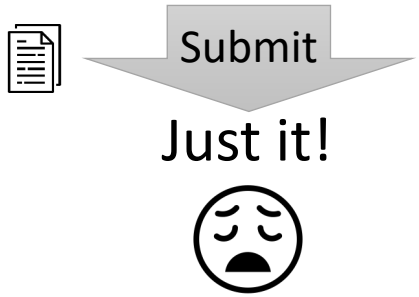
- The data management plan (DMP) is requested from funding agencies. However, it is hard to use for research support documents.

✓ **Unable** to compare the plan and current research status.

✓ **Unable** to support the research process based on the plan by the research office.

✓ **Unable** to utilize DMP by anyone, it is just a document to submit FA.

No	Data Title	Description	Owner	Format	Open Level	Confidentiality
1	Rabbit pupil response during flashlight stimulation.	Data was obtained from five albino rabbits. After 30 minutes of dark adaptation, …	Kazu YAMAJI, NII	Text	Level 4 (Public)	No

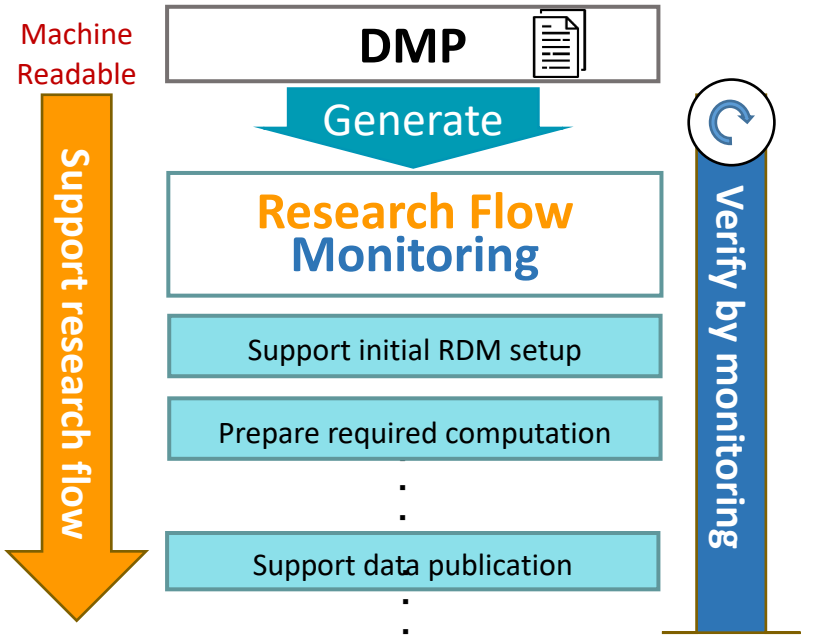


Solution

- Generate RDM environment in a semi-automatic manner based on DMP.

✓ **Able** to ensure RDM quality by the research flow generated usign DMP

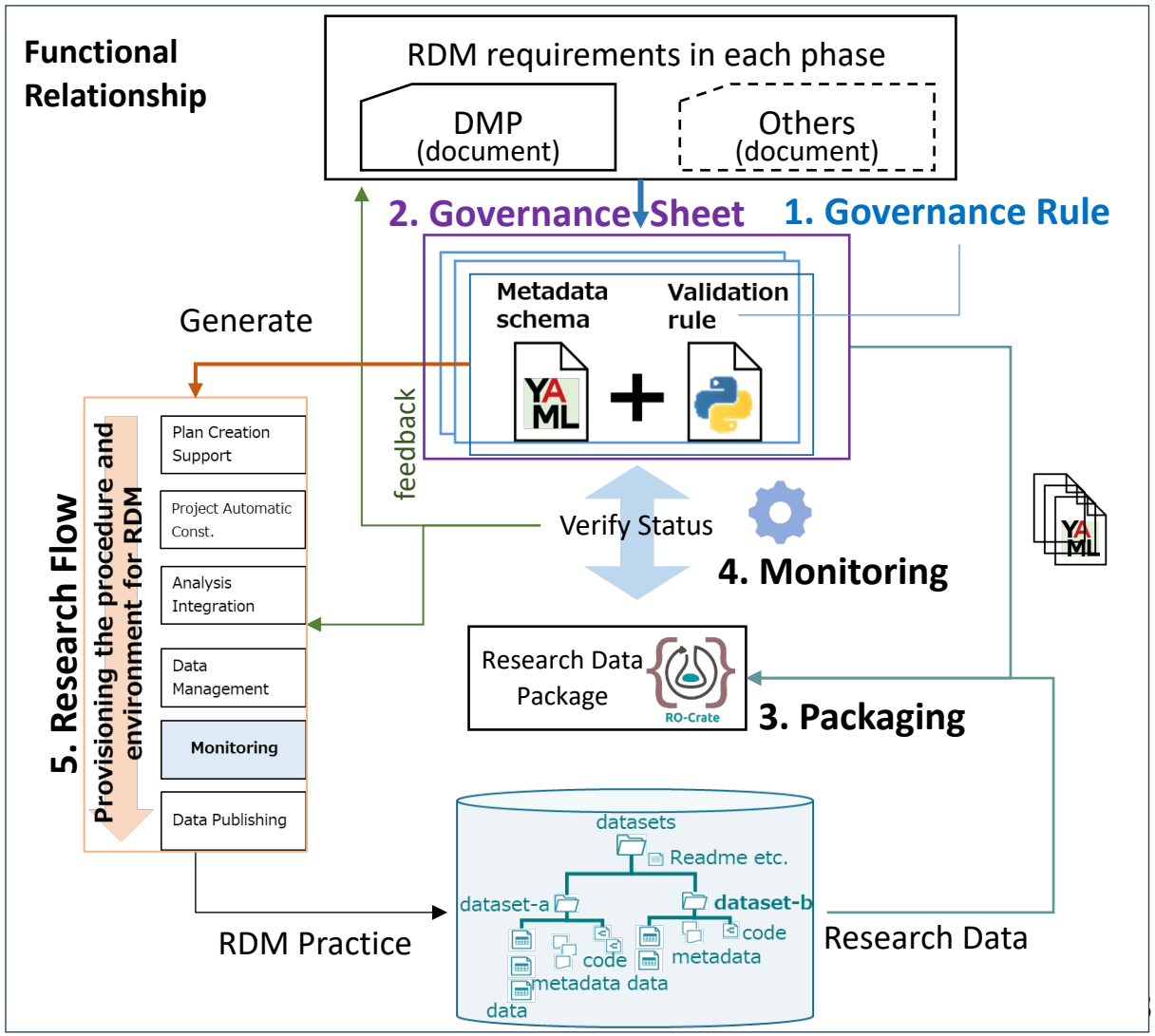
✓ **Able** to verify research output by monitoring the research process.



DG Functions supporting RDM

Support RDM by means of machine actionable manner based on DMP

1. **Governance Rule (Plan)**
   Define RDM requirements.
2. **Governance Sheet (Plan)**
   Configure RDM requirements in each project and its phase. (Set of governance rule)
3. **Packaging (Check)**
   Collect metadata related to the RDM requirements.
4. **Monitoring (Check/Action)**
   Check RDM statement whether it satisfies requirements defined in the governance sheet.
5. **Research Flow (Do/Action)**
   Provision research procedure and research environment according to the RDM requirements.

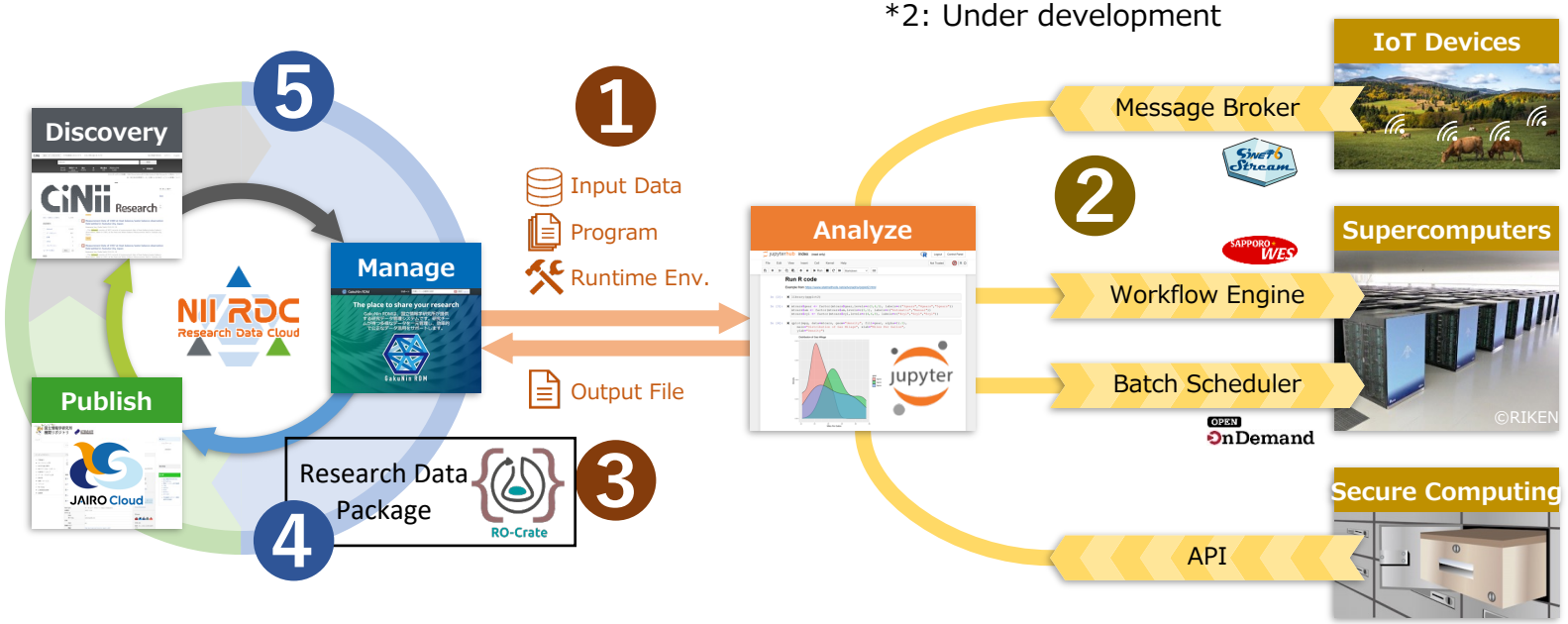


Research Reproducibility Functions

To promote reproducible data-driven science,
our Data Analysis Function supports:

1. Click and launch a JupyterLab environment on NII cloud*1. Start analyzing data in GakuNin RDM at ease.
2. Utilize various data sources and computing resources*2.
3. Share the results, the code and the runtime environment within the RDM project.
4. Publish the project as "reproducible code package".
5. Found and reused by another researcher.

*1: Other institutional/cloud computers can also be used
*2: Under development



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.

Management
Data Governance

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

Trust
Data Provenance

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

Interoperable
Data Curation

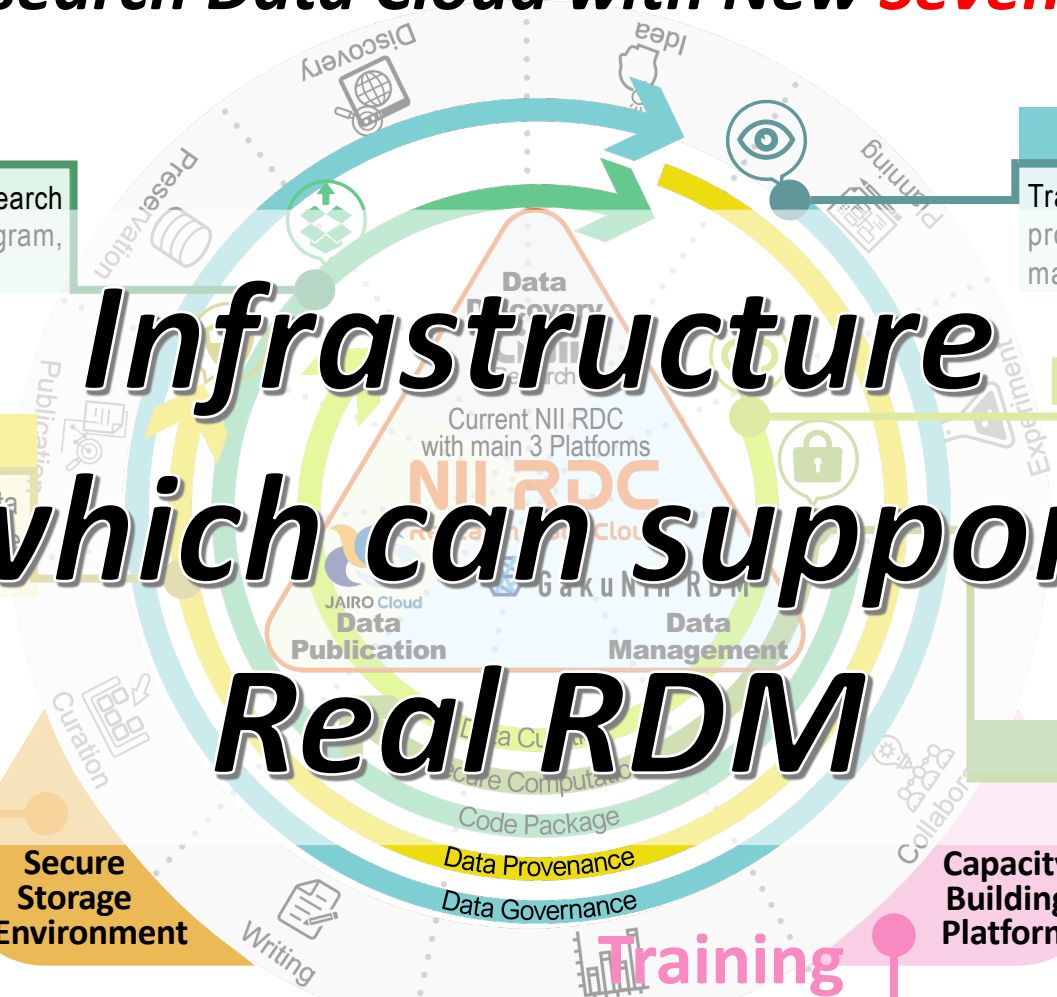
Contributes data reuse by developing an eco-system for data curation networks.

Safe
Secure Storage

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

Protection
Secure Computation

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



Infrastructure which can support Real RDM

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

RCOS