

Enabling Cross-Search across Social Science Data Archives in Japan —Initiative as part of National Endeavor to Establish Open Science Infrastructure

NATIONAL INSTITUTE OF INFORMATICS

- MIHO FUNAMORI, MASAHARU HAYASHI, KAZUTSUNA YAMAJI

RIKKYO UNIVERSITY


- MAKOTO ASAOKA, YUTAKA MAEDA (now at Kwansei Gakuin University)

THE UNIVERSITY OF TOKYO

- SATOSHI MIWA

IASSIST & CART 2018

May 31, 2018



Today's Talk

1. Status Quo and Issues of SSDAs in Japan
2. Policies and Platforms for RDM in Japan
3. System Configuration Options for Cross-searchable SSDAs in Japan
4. Pilot Study and Future Prospects for Cross-searchable SSDAs in Japan

1

Status Quo and Issues of SSDAs in Japan

Status Quo of Social Science Data Archives in Japan

□ Several small SSDAs existing

➤ Not inter-connected

- ✓ Not cross-searchable
- ✓ Not even connected organizationally

➤ Facing extinction

- ✓ Run by very small number of people
- ✓ Often times vague funding

Social Science Data Archives in Japan

□ General SSDAs

- SSJDA @ U Tokyo
- RUDA @ Rikkyo Univ
- SRDQ @ Osaka Univ
- SORD @ Sapporo Gakuin University <- stopped

□ Thematic SSDAs

- RIWQC-DA @ Japan Woman's Univ
- JPSC @ Institute for Research on Household Economics
- JILPT-DA @ Japan Institute for Labour Policy and Training
- JHPS/KHPS, JCPS, GEES @ Keio Univ
- JEDI @ Hyogo Univ of Teacher Education <- stopped

□ Governmental Survey Micro Data

- Hitotsubashi Univ
- KUMiC@ Kobe Univ
- Hosei University

Expectations

1. Make SSJDAs in Japan cross-searchable

- Make Japanese SS-related data **findable**.
- **Enhance visibility** of Japanese SS-related data domestically and internationally.

2. Provide stable data preservation platform

- **Data rescue**
- **Establish organizational procedures** for users to request data

2

Policies and Platforms for Research Data Management (RDM) in Japan

Policy Trends on RDM and Open Science in Japan

- June 2013
 - A joint statement by the G8 Science Ministers on making research data open
- March 2015
 - Cabinet Office, "Promoting Open Science in Japan"
- January 2016
 - "The 5th Science and Technology Basic Plan" ⇒ Open Science has been promoted in order to reinforce the intellectual infrastructure
- February 2016
 - Council for Science and Technology, "Promoting Open Access to Academic Information"
- July 2016
 - Science Council of Japan (SCJ), "Recommendations Concerning an Approach to Open Science that Will Contributes to Open Innovation"

Open Science Report from Japanese Cabinet Office (2015)

Promoting Open Science in Japan Opening up a new era for the advancement of science **Executive Summary**

Report by the Expert Panel on Open Science, based on Global Perspectives
Cabinet Office, Government of Japan

March 30, 2015

It is vital for Japan to participate in international discussions and to demonstrate a proactive approach to the promotion of open science. The Expert Panel on Open Science based on Global Perspectives has discussed various relevant issues of immediate importance for Japan. Based on these discussions, the Panel presented the guiding principles for promotion of open science in Japan.

I. The Importance of Open Science

"Open science" refers to a new approach to promoting innovation through knowledge creation in science and technology. This will be realized by facilitating access to and use of publicly funded research results such as scientific papers and their underlying data by the scientific community, industry and the general public. The concept of open science is spreading rapidly. At the G8 Summit held in June 2013, G8 Science Ministers issued a joint statement that endorsed the need for increasing access to publicly funded research, including peer-reviewed published research and research data. The statement triggered discussions in various forums worldwide

Research community, and to the decline of Japan's international competitiveness.

Japan should keep pace with the global advancement of open science in a collaborative yet also strategic manner, so that the value of Japan's latest research and development activities can lead to business activities at the next stage.

II. The Need to Promote Open Science

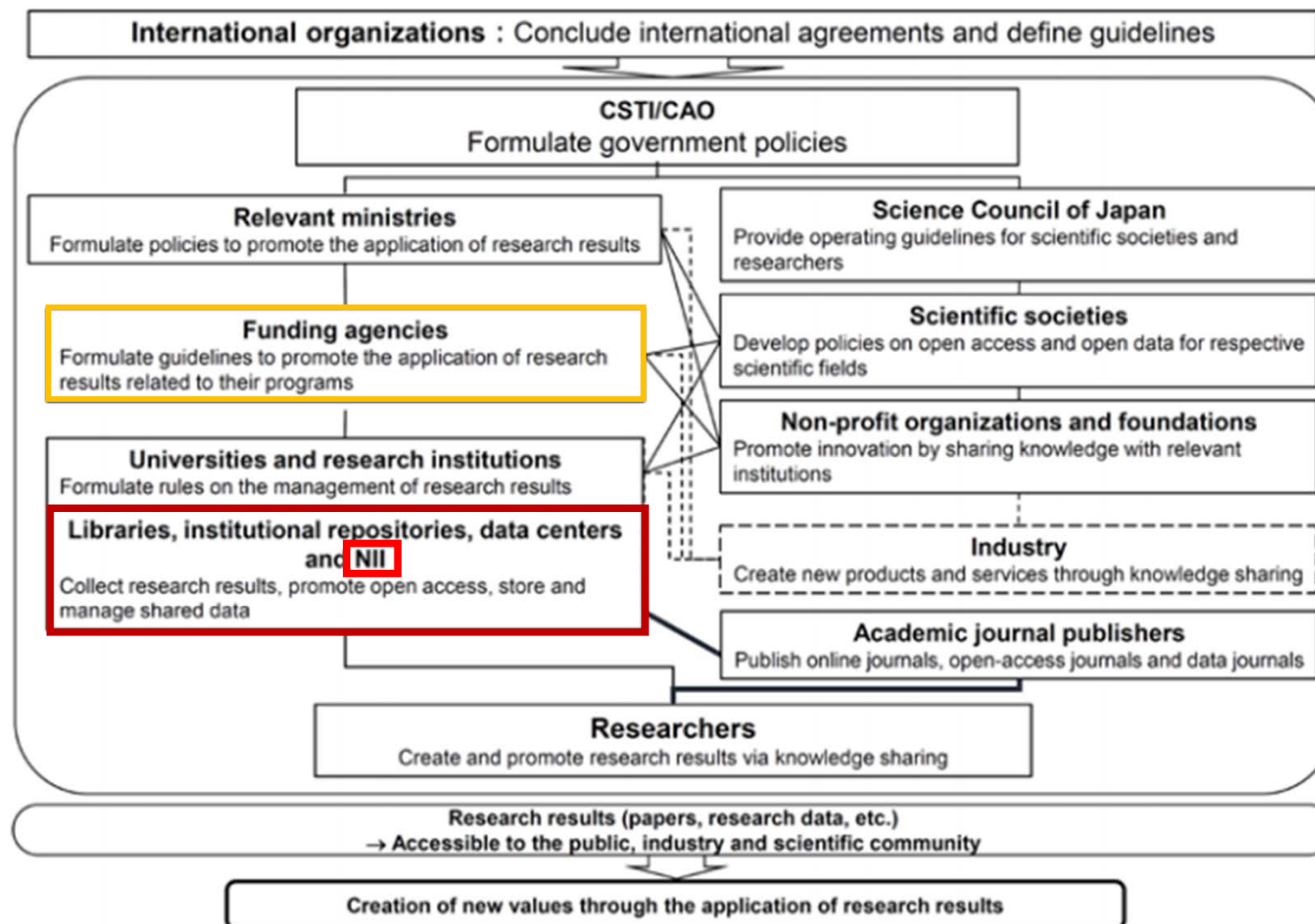
Open science may change scientific research. It will not replace traditional research methods, but will add new tools that help to advance science. It will make research results widely available in digital formats to all users including the scientific community, industry and the general public. This will enable additional value to be extracted from science and technology information, which will not only improve our knowledge, but will also reform innovation strategies.

For the scientific community, the acceleration of data-driven activities is expected to lead to new collaborations and to the prevalence of new research methods among researchers within the same research discipline and beyond. Industry and individuals are also expected to gain as they develop new products and services as a

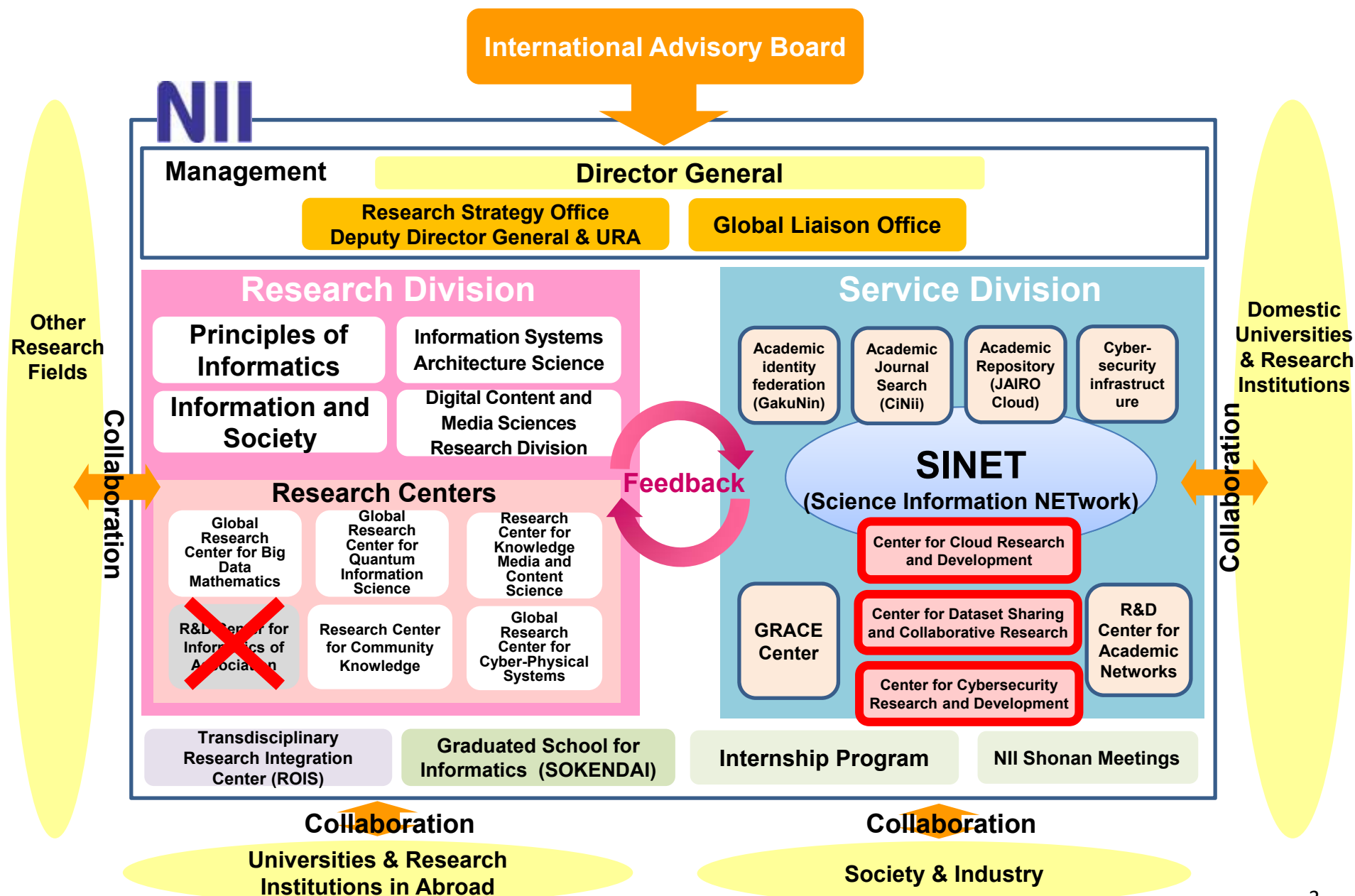
http://www8.cao.go.jp/cstp/sonota/openscience/150330_openscience_summary_en.pdf

Framework of the Open Science in Japan

Correlation diagram of policy making and implementation



NII's R&D Systems



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



Federation

- ◆ Collaborative enhancement of authentication between universities



Cloud

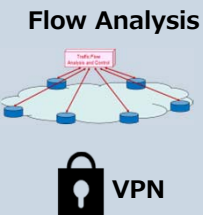
- ◆ Dramatic cost reduction and enhancement of research and education environment by tailored cloud services



GakuNin-Cloud
Direct Connection

Security

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

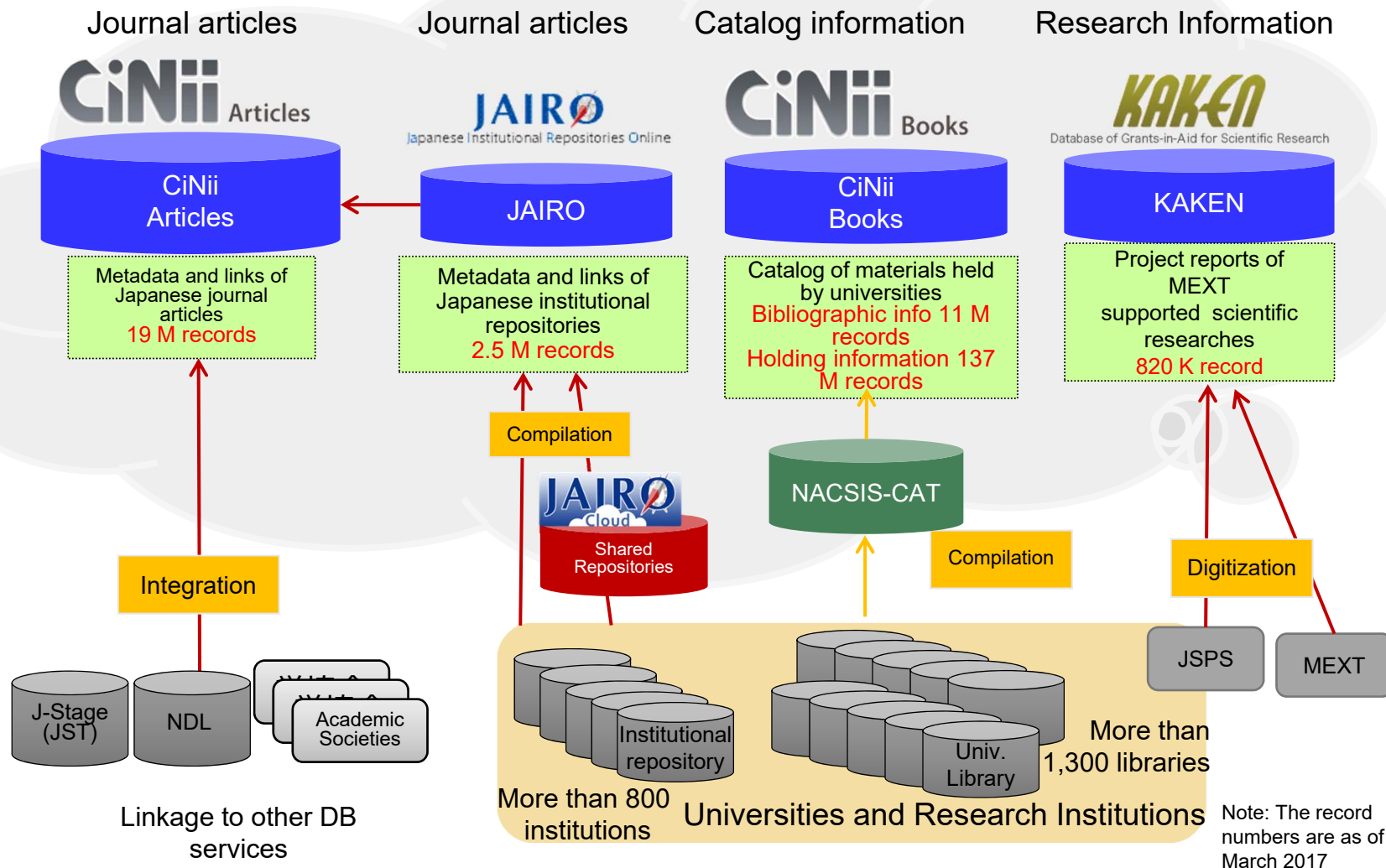


Network

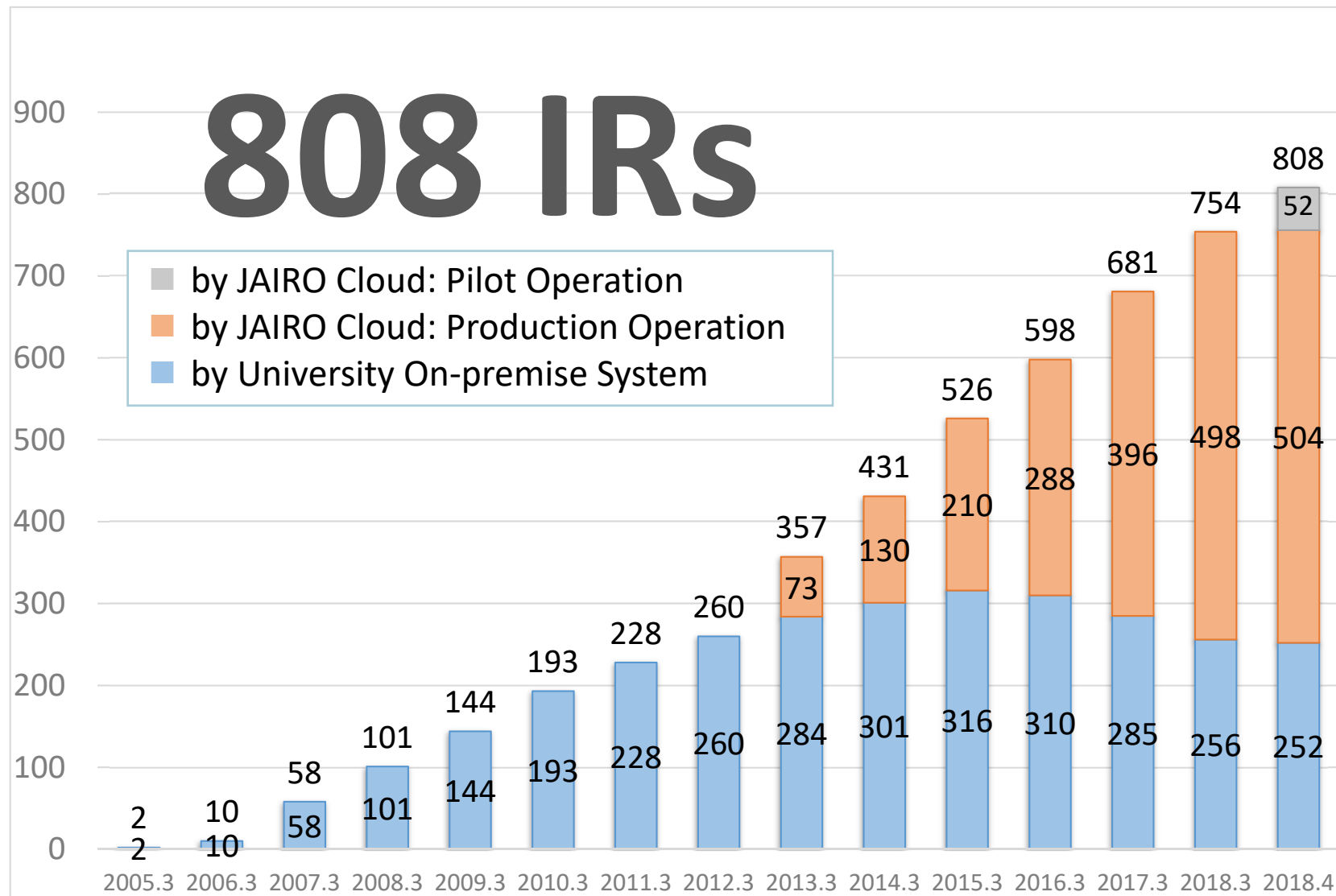
- ◆ Nationwide 100-Gbps backbone network and scalable network expansion
- ◆ High-speed direct international lines to USA, Europe, and Asia
- ◆ Introduction of new technologies such as SDN in response to user needs



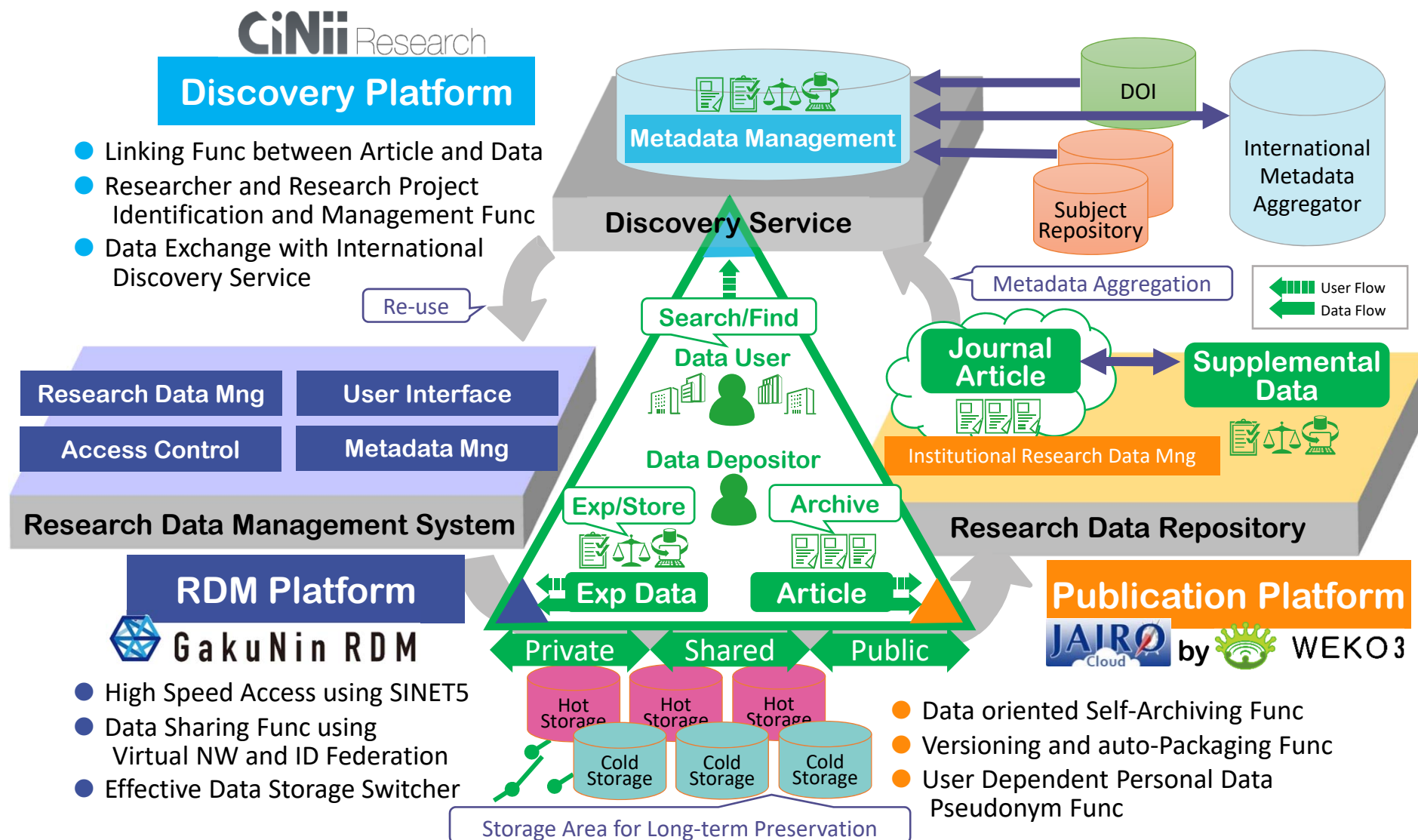
Scholarly Information Infrastructure



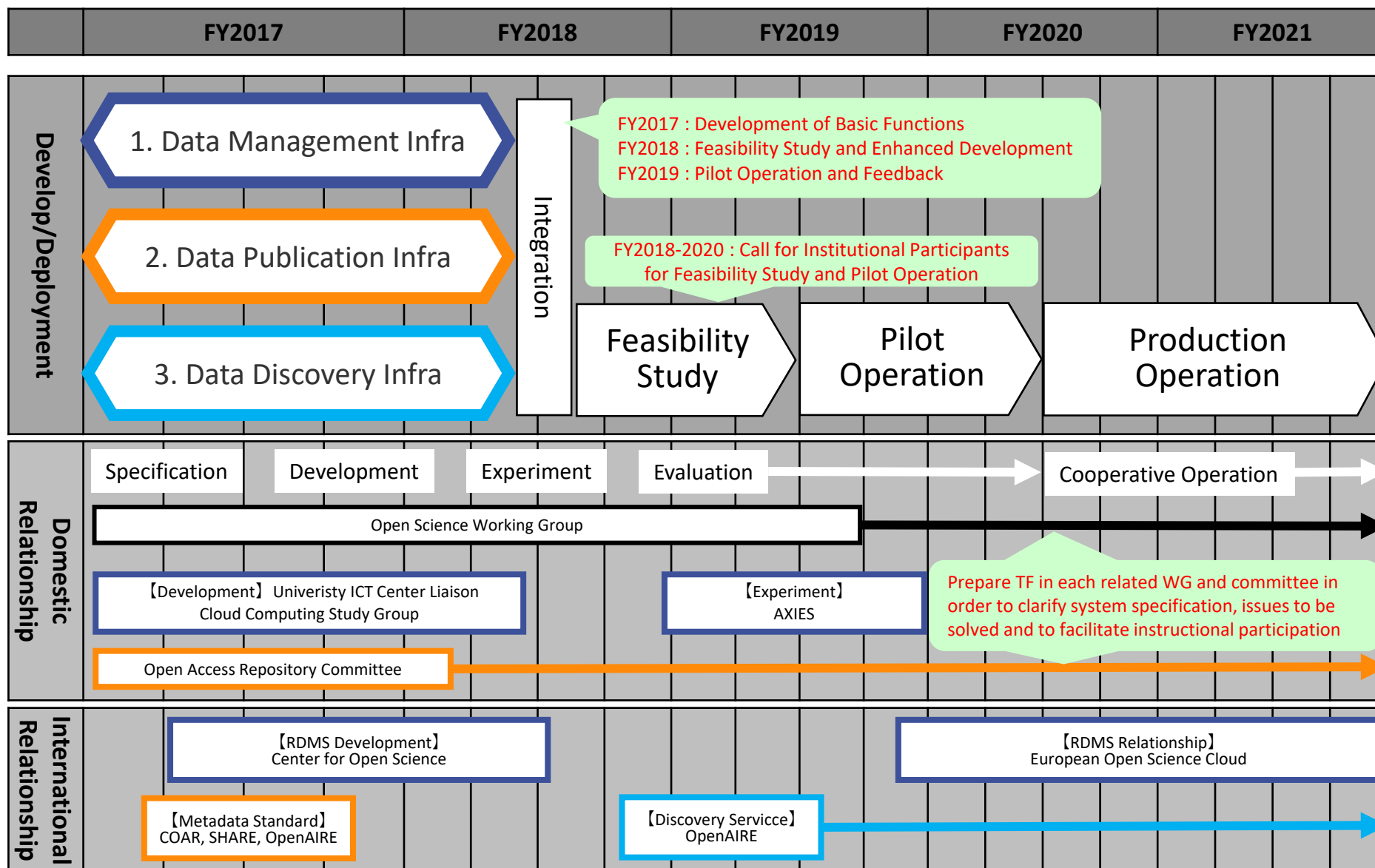
Number of Institutional Repositories in Japan



Research Data Infrastructure for Open Science



Planning



3

System Configuration Options for Cross-searchable SSDAs in Japan

Cross-searchable SSDAs in Japan

...Basic Concept (1)

- Use the NII Open Science Platform as data integrating mechanism.
 - However, as the OS platform is not yet available, use the existing platforms designed for research publications.
- SSDA Partners
 - Social Science Japan Data Archive (SSJDA)
 - Run by the University of Tokyo
 - Rikkyo University Data Archive (RUDA)
 - Run by the Rikkyo University

Cross-searchable SSJDA in Japan

...Basic Concept (2)

❑ Register only the metadata on the NII platforms while leaving the survey data at the original SSDA.

➤ Search results will redirect users to the original SSDA.

This will...

➤ Keep sensitive data at original SSDA.

➤ Utilize user consulting functions at SSDA.

Various Implementation Options

1. Institutional Repository Layer

- ✓ Where the metadata of SSDAs are going to be registered.

2. Discovery Layer

- ✓ Various academic information discovery services are available in Japan.

3. Metadata Schema

- ✓ What metadata schema to use for common SSDA.

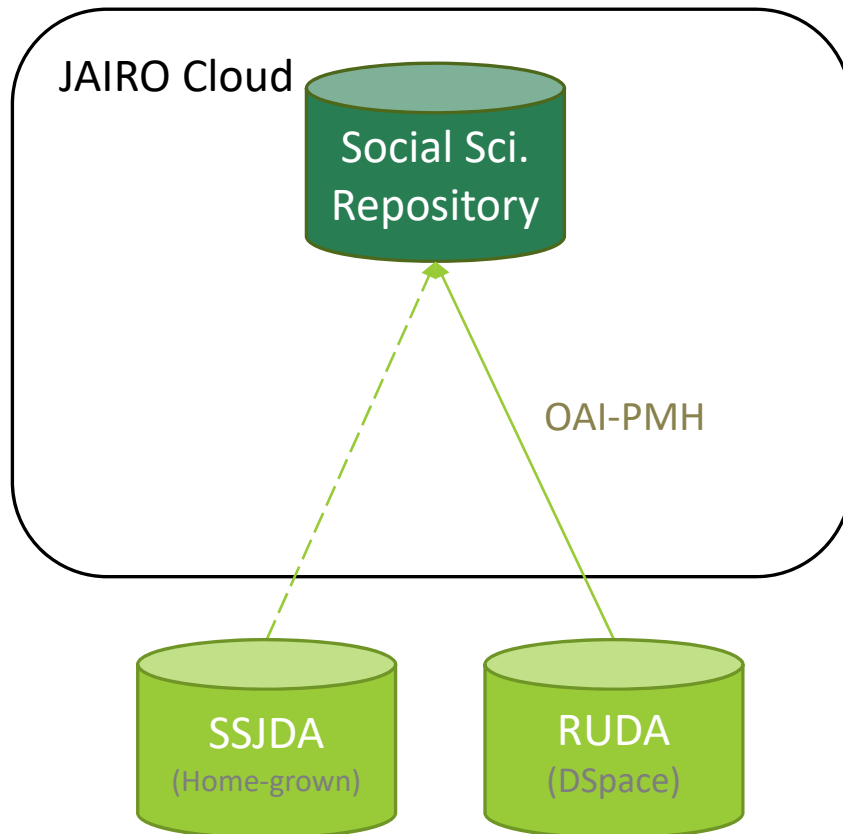
4. DOI Stewardship

- ✓ Various options for institutions to provide DOI stewardship.

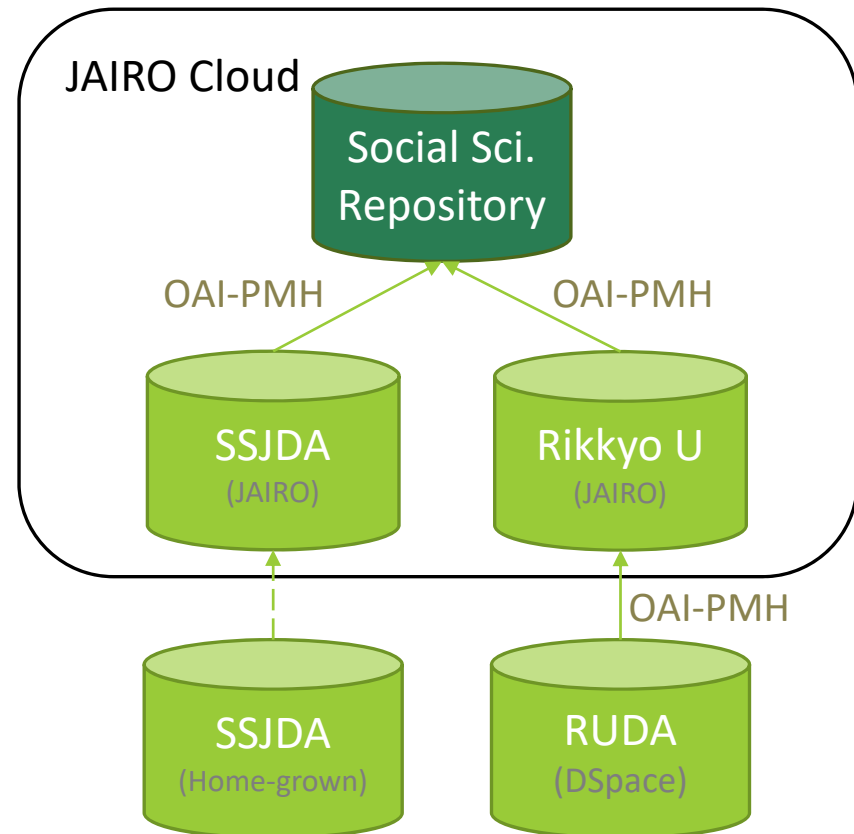
Various Implementation Options

...Institutional Repository Layer

A) Direct Register to SS Repository



B) Transferred Register to SS Repository



Various Implementation Options

...IR Layer...Pros and Cons

A) Direct Register

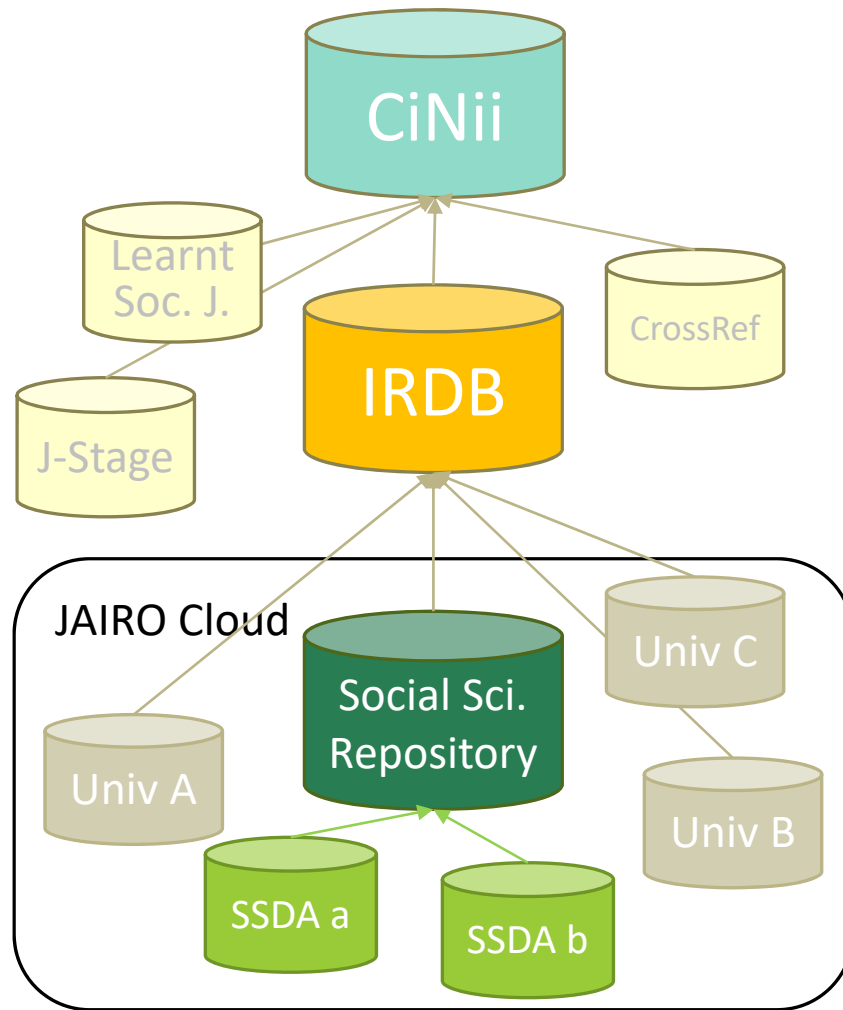
- **PRO**—Simple !
- **CONS**— Need administrative body for central repository in the long run.

B) Transferred Register

- **PRO 1**—Able to use stewardship of SSDA or university library at affiliated institution.
- **PRO 2**—Able to use OAI-PMH in the second stage for SSJDA.
- **CONS 1**—Administrative burden to maintain additional institutional repository whether managed by SSDA or university library.
- **CONS 2**—Additional fee for IR cloud service.

Various Implementation Options

...Three Discovery Layer



D) Cross-Academic
Resources discovery

C) Cross-IR discovery

B) SS-specified discovery

A) Single SSDA discovery

Various Implementation Options

...Discovery Layer...Pros and Cons (1)

A) Single SSDA discovery

- **PRO**—Good to promote the particular SSDA.
- **CONS**— Does not serve the cross-search purpose.

B) Social Science-specified discovery

- **PRO 1**—Searches only SSDA data and serves the SS academic community.
- **PRO 2**—Plan to provide refined search according to metadata schema used.
- **CONS**— Will not be known to people outside SS community.

Various Implementation Options

...Discovery Layer...Pros and Cons (2)

C) Cross-IR discovery

- **PRO**—Searches basically open accessed resources.
- **CONS 1**— Not very meaningful to search only IRs.
- **CONS 2**— IRDB search not well known by researchers.

D) Cross-Academic Resources discovery

- **PRO 1**—CiNii a well-known brand for academic search in Japan.
- **PRO 2**—Able to be used inter-disciplinary.
- **PRO 3**—Able to search for various academic resources (articles, books, data, etc.) in linked manner in future.
- **CONS 1**—Does not serve research data yet.
- **CONS 2**—Cannot refine search for SS-specific needs.

Metadata Schema

□ Implementation

- **JAIRO Cloud allows to use user-defined metadata schema.**
For the pilot study of cross-searchable SSDA, a metadata schema similar to DDI2.5 is used.
- **For harvesting, it uses junii2, a schema similar to Dublin Core.**
The metadata used at registering is mapped to junii2.
 - ✓ Junii2 will be upgraded to JPCOAR schema which is adjusted to research data use and not just for publication.

□ Issue

- **Existing SSDAs use usually self-defined metadata schema and are difficult to adjust to common standard.**
 - ✓ **Common metadata (fine)**—Need much effort to register all datasets.
 - ✓ **Common metadata (rough)**—Original information-level downgraded.
 - ✓ **Allow to use self-defined metadata**—Search accuracy goes down.

Adding DOIs

□ Implementation

- **Prefix**—Able to obtain from Japanese registration agency Japan Link Center (JaLC).
- **Suffix**—Able to obtain at JAIRO Cloud systematically.
- **Harvesting to IRDB and CiNII**—need to be an associate member of JaLC and apply for data provision.

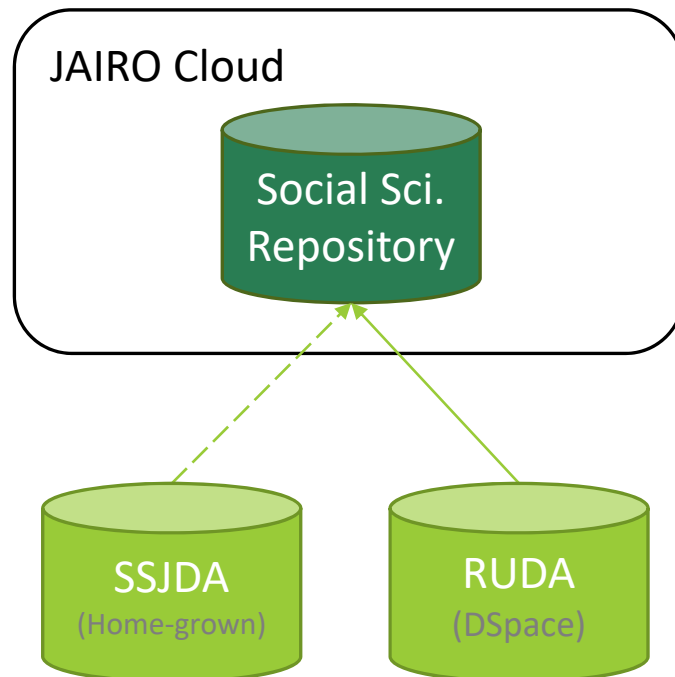
□ Issue

- **DOI can only be added to datasets which can assure long-term data provision.**
 - ✓ **Each SSDA**—Usually on weak grounds
 - ✓ **SSDA affiliated institution**—long lasting; however not sure if the institution will maintain SSDA assets forever.
 - ✓ **Cross-searchable SSDA**—Administrative body not determined. Also if existing, not able to assure its long-lasting existence.

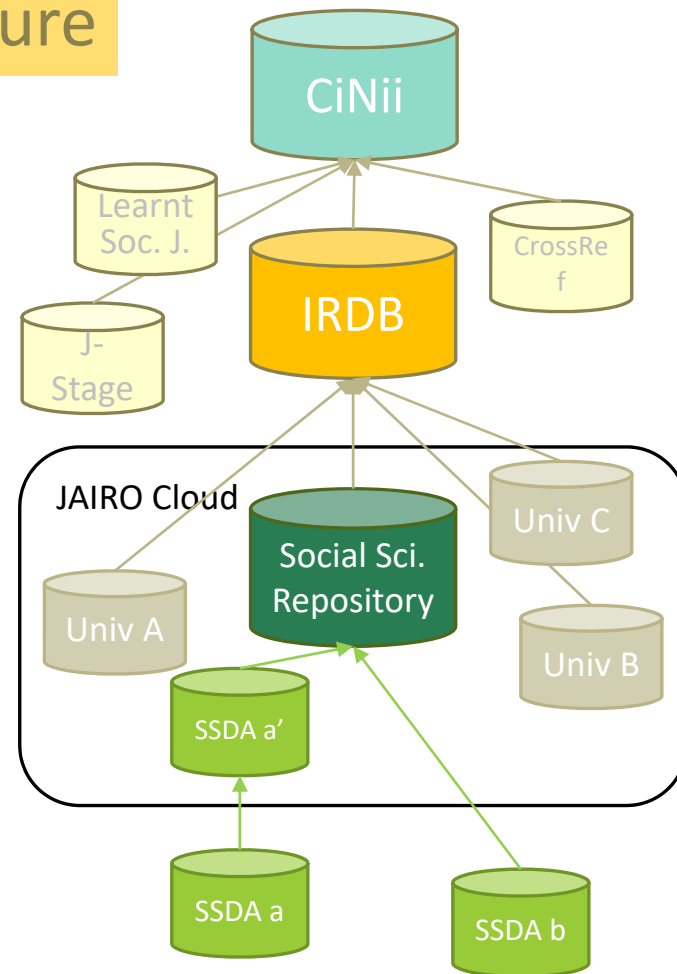
Cross-searchable SSDAs in Japan

...System Configuration

Pilot Study



Future



4

Pilot Study and Future Prospects for Cross-searchable SSDAs in Japan

Social Science Data Archives to post datasets at pilot study

□ SSJDA @ U Tokyo

- Roughly 1600 datasets in data archive
- Post 3 datasets for pilot study

□ RUDA @ Rikkyo University

- Roughly 50 datasets in data archive
- Post 10 datasets for pilot study

✕ This time, has taken the lead to formulate
metadata for common registering, similar to DDI 2.5

SSJDA @ U Tokyo



[Japanese](#)

Center for Social Research and Data Archives,
Institute of Social Science, The University of Tokyo

Data Search

Full Text :

Depositing Institution :

Topic (Click [here](#) for details) :

Unit of Analysis (Click [here](#) for details)

Survey Title :

Survey Year :

FROM: TO:

For Educational Purposes

Sort by:

Search

Reset

* Click on the survey title to display a description of the data.

<< < 1 2 3 4 5 6 7 >>> Total: 1308

List of Data

Survey No.	Survey Title	Depositor (Former Name)	Topic	Nesstar
0001	Labor Market Research among New Graduates (Junior High School), 1953	Institute of Social Science, The University of Tokyo	Employment/Labor	--
0002	Fact-finding Survey on Wage Setting among Small and Medium-sized Enterprises, 1995	Research Institute for Advancement of Living Standards	Employment/Labor	--
0003	Survey on Young Workers' Professional and Academic History, 1995	Research Institute for Advancement of Living Standards	Employment/Labor Education/Learning	--
0004	Questionnaire Survey on Daily Life among Elementary and Junior High School Students, 1995	Research Institute for Advancement of Living Standards	Society/Culture Education/Learning	--
0005	Questionnaire Survey on Volunteer Activities, 1995	Research Institute for Advancement of Living Standards	Society/Culture Security/Crime/Disaster	--

RUDA @ Rikkyo University



RUDA Home / Search

BROWSE

Find Data

- ▶ Geographic Coverage
- ▶ Date of Distribution
- ▶ Authoring Entity
- ▶ Title
- ▶ Subject Terms
- ▶ Discipline

MY ACCOUNT

- ▶ Login
- ▶ Register

DISCOVER

Survey Region

- ▶ Tokyo/Japan (28)

Search

Find Data

Go

Show Advanced Filters

Now showing items 1-10 of 125



Communities or Collections matching your query

01 日本 | 01 Japan [7]
地域を特定しない調査

Items matching your query

No Thumbnail

NI 日本の国際化についての世論調査

IS 石田 淳 (2017-02-12)

現代日本のナショナル・アイデンティティの様態をプール・カテゴリー分析（石田 2016）によって把握するために、国籍・在住・血統・言語の4条件からなる16通りのヴィネットを構成し、それぞれについて「日本人」であるかどうかの判断を求めた。同時に、ナショナル・アイデンティティ、ナショナリズム関連の質問、社会経済的地位に関する質問も行った。

Registered survey image

WEKO [Edit](#)

[Top](#)
[Item Registration](#)
[Workflow](#)
[Ranking](#)

[Detail](#)
☒ Full Text
 ☐ Simple

Language
 English ▼

Index Tree

- [-] RUDA
 - 0001-0005
 - [-] 0006-0010
 - 0006
 - 0007
 - 0008
 - 0009
 - 0010
- INDEX1
- INDEX2
- [-] SSJDA
 - [-] 0001-0005
 - 0001
 - 0002
 - 0003

Index
[RUDA](#) ▶ [0006-0010](#) ▶ [0006](#)
 Permalink : <http://id.nii.ac.jp/1605/00000030/>

2007 GMFS - 10 City Survey "Quality of Life Survey"
[Show Usage Statistics](#)

File / Name	License
調査票 (DL可) (254.4Kb)	
調査票 (DL可) (254.4Kb) (260.57KB) [0 downloads]	

Item Type	調査概要_DDI
Language	English
Keywords	Quality of Life, Happiness, Competitiveness
Survey Number	0006 Community life
Publications	The Global Metropolitan Forum of Seoul, 2008, The Second International Conference of the Global Metropolitan Forum of Seoul.
Condition	社会情報教育研究センターの承認による
Bibliographic Citation	The Global Metropolitan Forum of Seoul, 2007, 2007 GMFS - 10 City Survey (配布者: 立教大学社会情報教育研究センター, 公開日: 2010年10月1日) .
URL	https://ruda.rikkyo.ac.jp/dspace/handle/10577/92
Public/Private Status	Public Change to Private

[← Return](#)
[Edit](#)
[Delete](#)

WEKO

Top Ranking

ruda

[Detail](#) ☒ Full Text ☐ Simple

	Title ▾	<input type="text"/>	
AND	Author or Author IC ▾	<input type="text"/>	
AND	Resource Type ▾ list	<input type="text"/>	
AND	Date of Publication ▾	<input type="text"/> year - <input type="text"/> year	
AND	Index ▾ list	<input type="text"/>	
AND	WEKO Author ID ▾	<input type="text"/>	

[More Input Row](#)

Language

English ▾

Index Tree

- RUDA
- SSJDA

Search Results 1 - 13 of 13 items

export checked items ▾

Sort PubYear (DESC) ▾ Num of Items 20 ▾

[東京版総合社会調査「高齢・少子社会における都市居住と家族・親族関係に関する調査」](#)

0001 , 立教大学社会情報教育研究センター

[2007 GMFS - 10 City Survey "Quality of Life Survey"](#)

0006 , Rikkyo University Center for Statistics and Information

Issue 1...Administrative body for cross-searchable SSDA

□ Issue

- Need for a body which would oversee the cross-searchable SSDA.

□ Needed attributes

- Sustainable body
- Setting the standards for cross-searchable SSDA and maintaining it to certain degree
- Assisting other SSDAs to register their datasets

□ Options

- Steering Committee by member SSDAs
- One dedicated institute
 - ✓ a. new body, b. one SSDA taking up the role

Issue 2...Metadata Schema for cross-searchable SSDA

□ Issue

- What metadata schema should be used?

□ Facts

- Most SSDAs use home-grown metadata schema.
- To establish cross-searchable SSDA, a common metadata schema is desired.
- DDI, which would be an option, is not well-known in Japan.

□ Options

- DDI-like metadata
- Dublin Core-like metadata
- Home-grown metadata for each SSDA

Issue 3...Contents quality standardization for cross-searchable SSDA

□ Issue

- How should the contents quality be standardized?

□ Facts

- Even if same metadata label is used, different SSDAs have different arrangements in the usage.
 - ✓ Metadata “Key Word”—Free KWs versus Choice from categories

□ Options

- Use the original SSDA’s contents—Easy to implement. However, not useful for cross-search
- Standardizing to common quality—Data registering burden. However, refined search possible.

Other Issues for cross-searchable SSDA

□ Who is going to register the datasets?

- For established SSDAs, need efforts to adjust datasets quality to common standard but lack of manpower.
- For long tail datasets, if we allow self-archive the quality would not be sustained.

□ How do we adopt datasets of SSDAs which have stopped providing service?

- How do we transfer the rights of the original data provider? With whom do you negotiate?
- Who is going to do the work for registering datasets?
Who is going to pay for the work?

Future prospects

□ A new move

- A report by the Science Council Japan: “For the promotion of scholarship—A recommendation for humanities and social sciences”
- Funding from JSPS: “Humanities and Social Sciences Data Infrastructure Establishing Program,” (2018-)

□ Future steps

- Hope for the steering committee of JSPS program to be able to take leadership, have various SSDAs to join discussion for future vision.
- This bottom-up pilot study will serve to materialize the initiative of the JSPS program.