



The Japanese academic dataset integration
based on PID and text processing.

2023/06/14

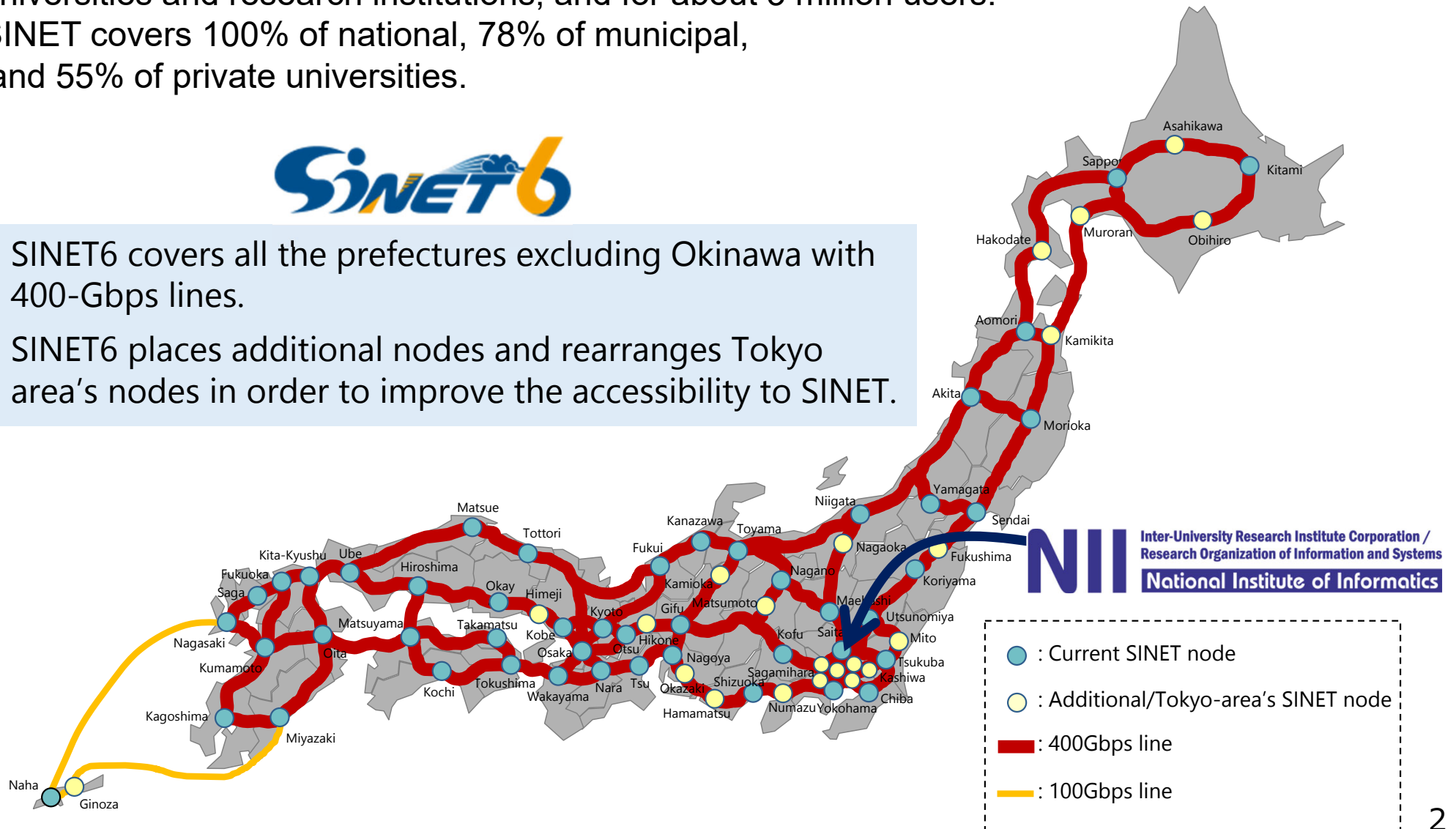
National Institute of Informatics (NII), Japan
Jun-ichi Onami

Background: NII provides the Japanese academic network

- NII is **Japan's only national institute of informatics** to create future value.
- NII also provides SINET6, which is a Japanese academic backbone network for nearly 1000 universities and research institutions, and for about 3 million users.
- SINET covers 100% of national, 78% of municipal, and 55% of private universities.

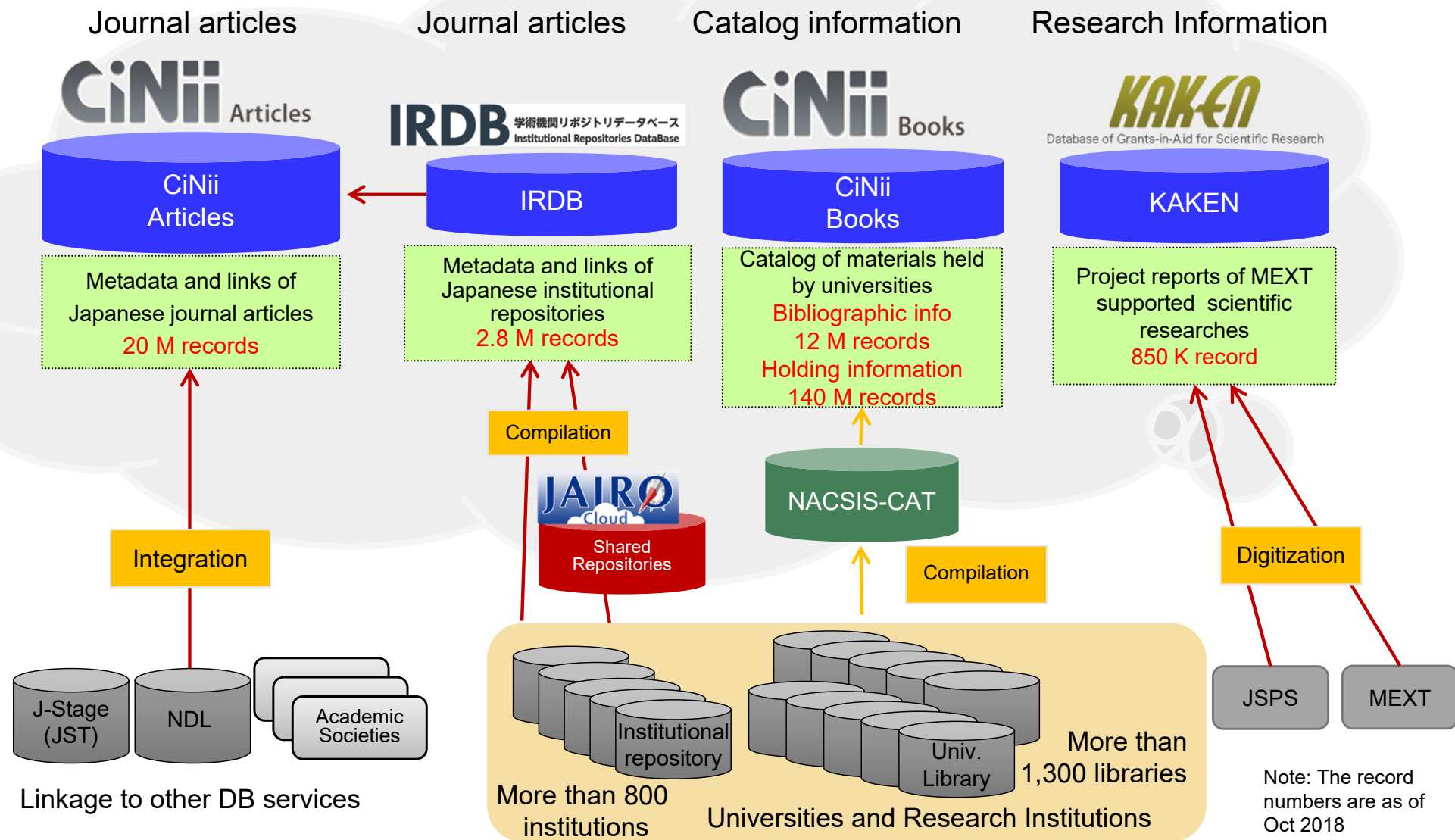


- SINET6 covers all the prefectures excluding Okinawa with 400-Gbps lines.
- SINET6 places additional nodes and rearranges Tokyo area's nodes in order to improve the accessibility to SINET.

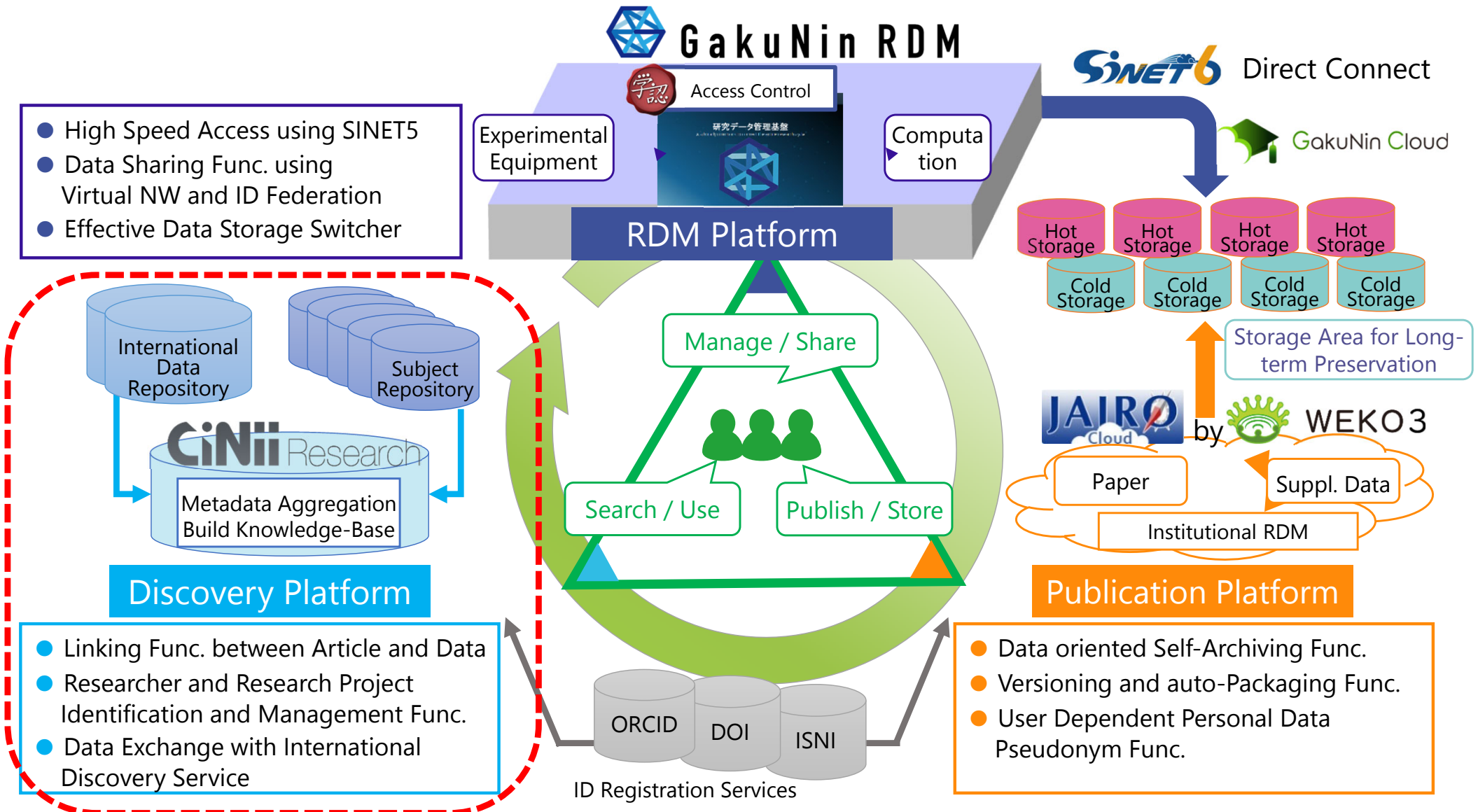


Japanese Scholarly Information Infrastructures

In spite of its huge amount of dataset, Japanese Scholarly Information Infrastructure is dispersed to many websites.



National Institute of Informatics started to provide a discovery platform from NII Research Data Cloud (NII RDC)



2017 Start Development ⇨ 2021: Start Operation

Discovery platform

CiNii Research

In April 2021, we published a new Japanese academic discovery service, CiNii Research.

Articles

Differential Diagnosis of COVID-19: Importance of Measuring Blood Lymphocytes, Serum Electrolytes, and Olfactory and Taste Functions
Nakanishi Hiroki, Suzuki Motohiko, Maeda Hiroyoshi, Nakamura Yoshitaka, Ikegami Yosuke, Takenaka Yuya, Mori Yusuke, Hasuo Takahiro, Hasegawa Chihiro The Tohoku Journal of Experimental Medicine 252 (2), 109-119, 2020

...However, clinical features, which can differentiate COVID-19 from non-COVID-19, are not clear. We therefore examined the key clinical features of COVID-19 and non-COVID-19 patients...

DOI PubMed Web Site ほか2件 参考文献19件

Demand and supply of invasive and non-invasive ventilators at the peak of the COVID-19 outbreak in Okinawa
Kuniya, Toshikazu, Nakama, Yasunori, Tokunaga, Masahito, et al. Journal of Infection Management and Control 101, 2020-07

...of critically ill patients with COVID-19...

機関リポジトリ CiNii

An equitable approach is necessary to win a war against the global COVID-19 pandemic
MITSUMORI Yaeoko International Journal of Japan Association for Management Systems 12 (1), 111-116, 2020

...<p>There seems to be no end to the current COVID-19 pandemic...

DOI Web Site 参考文献6件

Reference Information

<https://cir.nii.ac.jp/?lang=en>

CiNii Articles, Data Books Dissertations Login 日本語

[Updated on Apr. 18] Integration of CiNii Articles into CiNii Research

CiNii Research

Free word Search

All Data Articles Books Dissertations Projects Advanced Search

About CiNii CiNii Research Incorporated Databases Terms of Use Contact NII Home

Help CiNii Research CiNii Books CiNii Dissertations Metadata and API

News Updated RSS Twitter

NII Scholarly Services KAKEN IRDB NII-REO

Bibliographics

山中伸弥先生に、人生とiPS細胞について聞いてみた
山中, 伸弥, 緑, 慎也 講談社 2012

Thumbnail image by OpenBD API

平尾誠二と山中伸弥「最後の一年」
山中, 伸弥, 平尾, 誠二, 平尾, 恵子 講談社 2017

Dissertation articles

Noninvasive SAR measurement technique and antenna for biological assessment in body-centric wireless communication
Rasyidah Hanan Binti Mohd 東京農工大学 博士 (工学) 甲博工第1203号 2021-09-15

IR NDL Digital Collections CiNii

Characterizations and mechanisms of shallow lakes
Zhou, Yiwen 東京農工大学 博士 (工学) 甲博工第1203号 2021-09-15

IR NDL Digital Collections CiNii

Dissertation information

Research projects

Establishment of removal method of undifferentiated human iPS cells contained in differentiated cells.
NAKASHIMA YOSHIKI University of the Ryukyus 2016-08-26 - 2018-03-31 (KAKENHI)

Polymorph control of ...
Yamanaka Shinya Muraoka ...

Decadal variability of mode waters associated with that of the Kuroshio Extension and mesoscale eddies
OKA Eitarou The University of Tokyo 2013-04-01 - 2017-03-31 (KAKENHI)

Funding information

Research data

Development of a Model for the Spread of Nosocomial Infection Outbreaks Using COVID-19 Data
Ueda, Ryuchiro, Goto, Ayato, Kita, Ryuki, Ogasawara, Katsuhiko Healthcare 10 (3), 471-, 2022-03-03

HANDLE

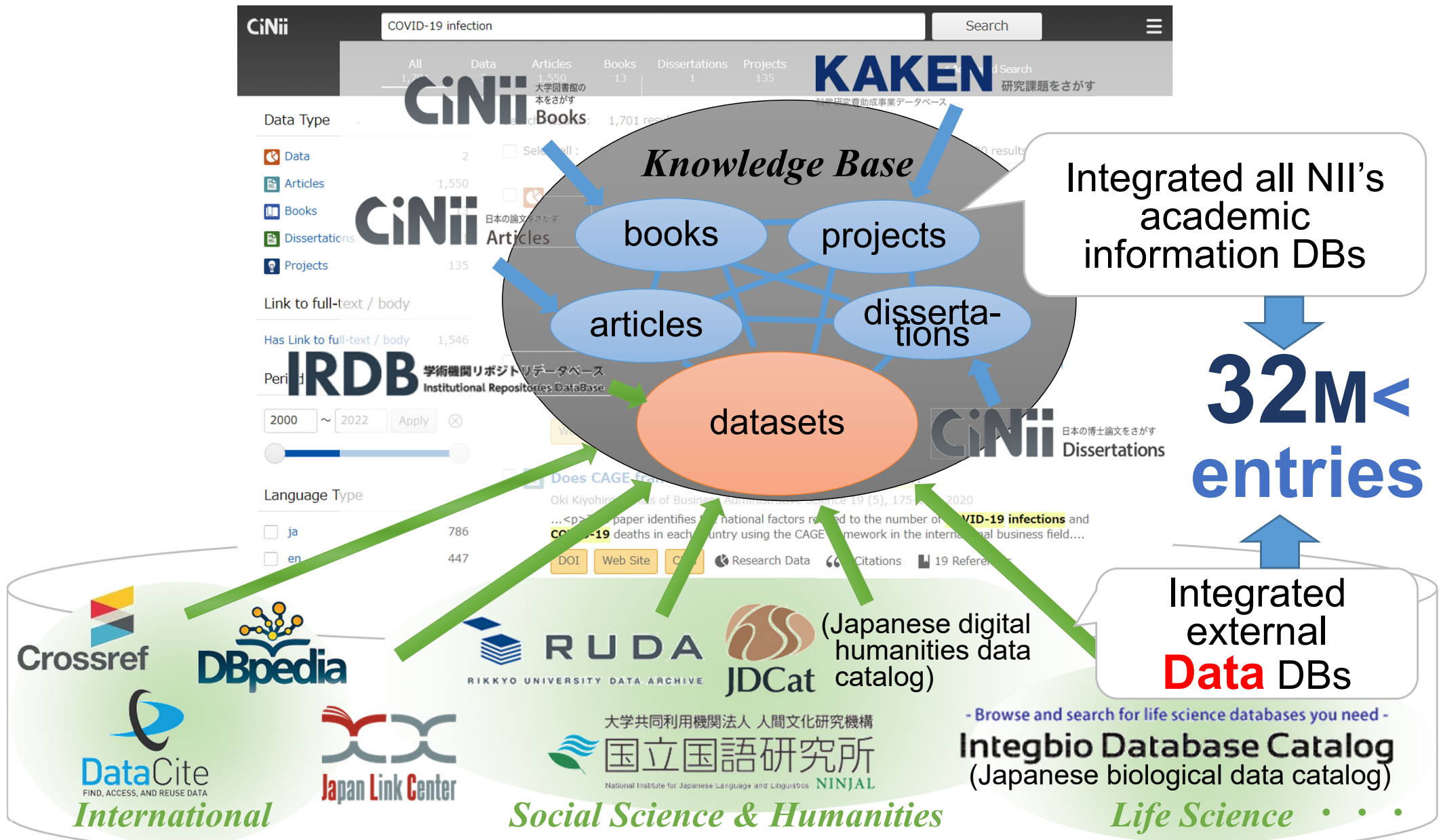
Data used in the regression analyses to identify the national factors related to the number of COVID-19 infections and COVID-19 deaths in each country and all results of the statistical analyses
Oki Kyojiro 2021-09-17 (Last modified at 2021-09-17)

...This file contains the dataset and the results related to the number of COVID-19 infections...

DOI DOI DOI

Access information for data resources

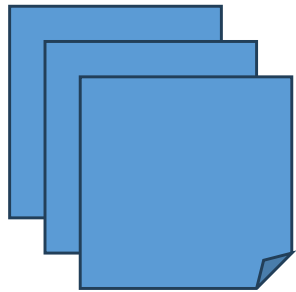
CiNii Research integrates various research outputs



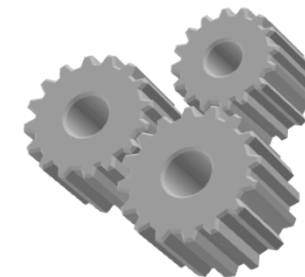
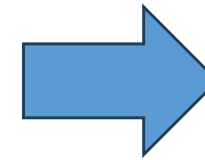
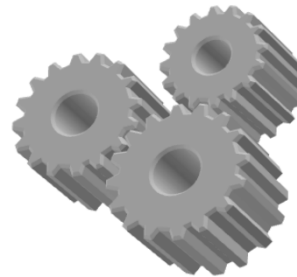
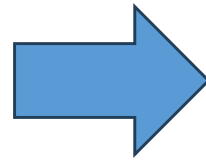
3 steps of data processing method

(Data process 1)
Integration using PID

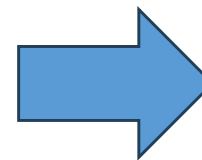
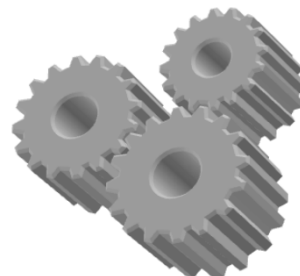
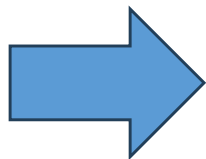
(Data process 2)
Integration with using text
matching and ID mapping



Collection of academic
data resources



(Data process 3) Connecting
instances by using citation
information and parent-child
information.

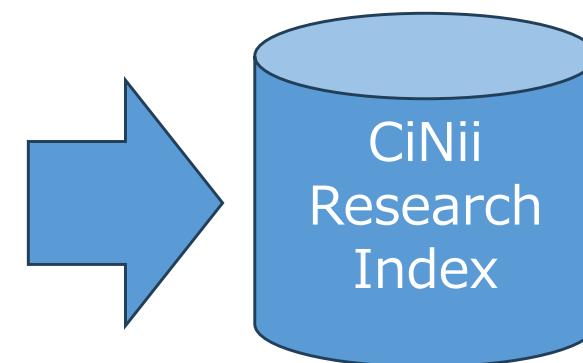


Fitting to Web UI

Integration steps of CiNii Research

(Data process 1) Integration using PID such as DOI of the instance.

	Number	ID		Number	ID
Projects	1	CRID	Researchers	1	CRID
	2	ERAD		2	AID
	3	KAKEN		3	CINII_AUTHOR_ID
	4	URI		4	ERAD
Research outputs	1	CRID		5	ERAD_AS_WRITTEN
	2	ARK		6	GRID
	3	ARXIV		7	ISNI
	4	BIBCODE		8	J_GLOBAL
	5	CODEN		9	KAKENHI
	6	COI		10	NRID
	7	DOI		11	ORCID
	8	DOI_AS_WRITTEN		12	ORCID_AS_WRITTEN
	9	EAN13		13	RESEARCHMAP
	10	EISSN		14	RINGGOLD
	11	HANDLE		15	ROR
	12	HDL		16	URI
	13	IGSN		17	VIAF
	14	ISBN		18	NDL_AUTHOR_ID
	15	ISBN_AS_WRITTEN		19	NDL_AUTHOR_ID_AS_WRITTEN
	16	ISSN	Organizations	1	AID
	17	ISTC		2	DOI
	18	JST		3	DOI_AS_WRITTEN
	19	J_GLOBAL		4	ERAD
	20	LCCN		5	GRID
	21	LISSN		6	ISNI
	22	LOCAL		7	JISX0408
	23	LSID		8	JSPS
	24	MRID		9	JST
	25	NAID		10	KAKENHI
	26	NCID		11	MEXT
	27	NDL_BIB_ID		12	NACISIS_CAT_KID
	28	PISSN		13	NII
	29	PMCID		14	NRID
	30	PMID		15	ORCID
	31	PURL		16	RINGGOLD
	32	SCOPUS		17	ROR
	33	UPC		18	VIAF
	34	URI		19	ERAD_INSTITUTION_CODE
	35	W3ID	20	ERAD_DEPARTMENT_CODE	
	36	WOS			
	37	CODEN			
	38	NDL_CALL_NUMBER			
	39	NDLJP			



All IDs were fitted to JPCOAR schema standard and indexed in CiNii Research.

Yellow highlights: PID

Integration steps of CiNii Research

(Data process 2) Integration with using text matching and ID mapping of metadata after text normalization.

metadata	JPCOAR schema target
title	* content.title.notation.text
author	* authors.text * containing the first and last name * matching first author's name
journal title	* content.publication.publicationTitle.text
publication year	* content.publication.issued
journal ISSN	* content.publication.publicationIdentifier
DOI	* jpcoar:identifier

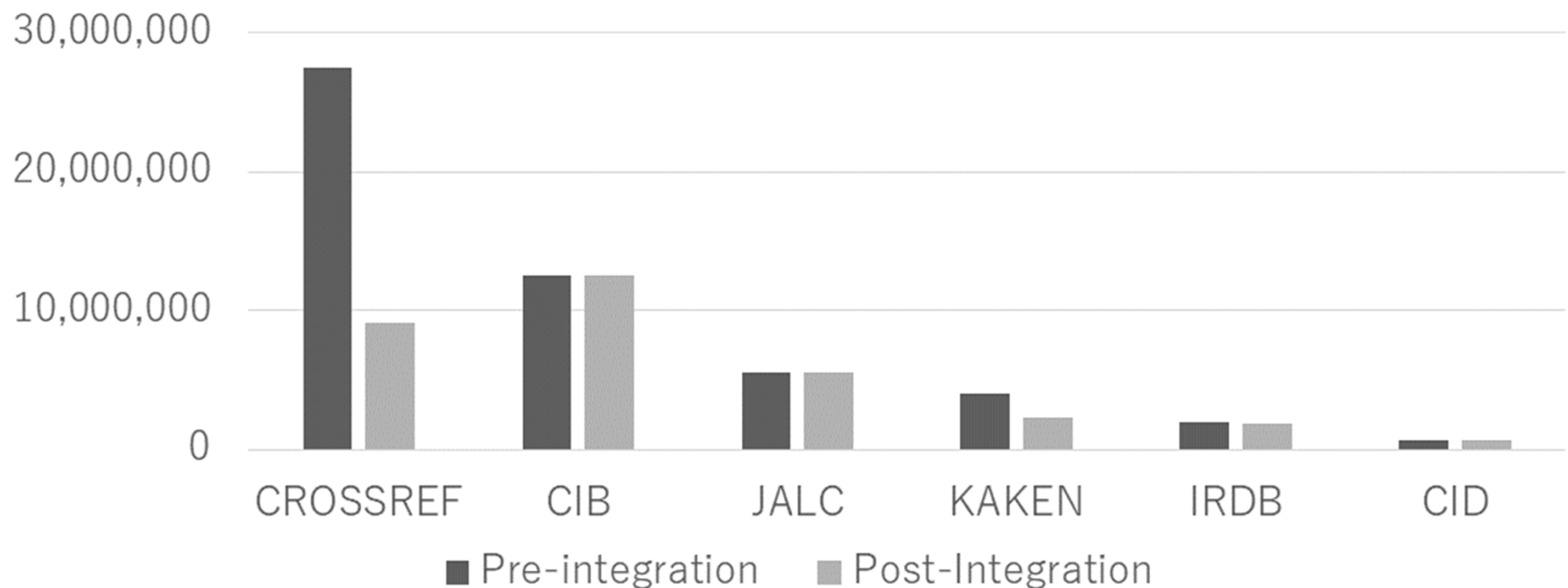
Text normalization process contains the removal step of blank character and symbol character, and the changing step of uppercase letter to lower case letter.

If more than two metadata were exactly matched, integration executed.

(Data process 3) Connecting instances by using citation information and parent-child information.

The effect of the integration steps

we confirmed the reduction of thirty-nine percent of instances by integration



The data obtained by CrossRef was significantly reduced, suggesting that the result of interoperable, which is recommended as "I" in the FAIR principle, is adversely affected by redundancy, which is a problem in the field of research data quality.

Relation datasets in CiNii Research

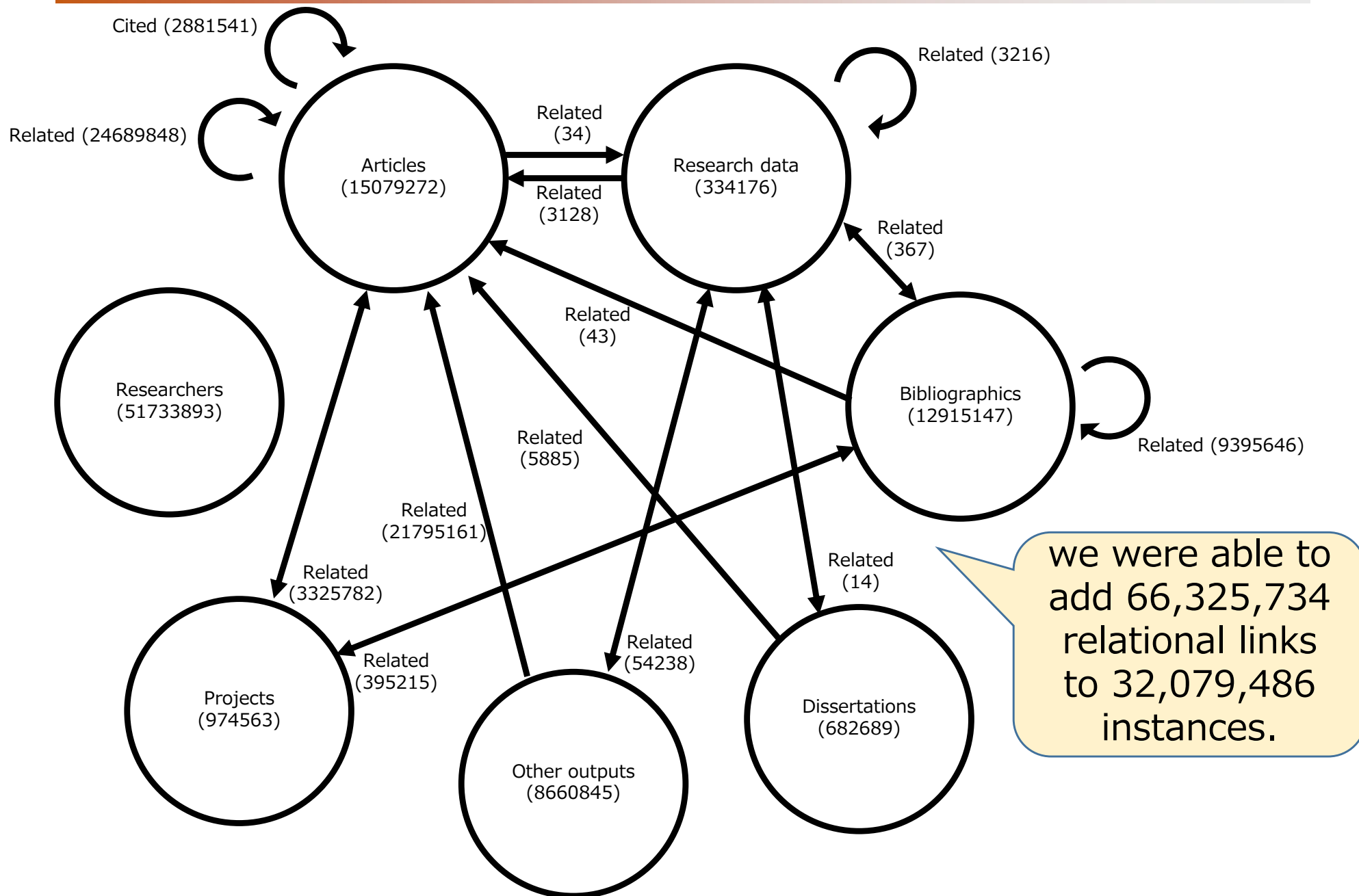
We added the following relation datasets from four data resources.

Relation	Inverted relation	Source
cites	isCitedBy	DataCite
isCitedBy	cites	DataCite
compiles	isCompiledBy	DataCite
isCompiledBy	compiles	DataCite
continues	isContinuedBy	DataCite
isContinuedBy	continues	DataCite
describes	isDescribedBy	DataCite
isDescribedBy	describes	DataCite
documents	isDocumentedBy	DataCite
isDocumentedBy	documents	DataCite
hasFormat	isFormatOf	jpcoar
isFormatOf	hasFormat	jpcoar
hasMetadata	isMetadataFor	DataCite
isMetadataFor	hasMetadata	DataCite
hasPart	isPartOf	jpcoar
isPartOf	hasPart	jpcoar
hasVersion	isVersionOf	jpcoar
isVersionOf	hasVersion	jpcoar
isDerivedFrom	isSourceOf	jpcoar
isSourceOf	isDerivedFrom	jpcoar
isIdenticalTo	isIdenticalTo	jpcoar
isNewVersionOf	isPreviousVersionOf	DataCite
isPreviousVersionOf	isNewVersionOf	DataCite
isObsoletedBy	obsoletes	DataCite

Relation	Inverted relation	Source
obsoletes	isObsoletedBy	DataCite
isOriginalFormOf	isVariantFormOf	DataCite
isVariantFormOf	isOriginalFormOf	DataCite
isReferencedBy	references	jpcoar
references	isReferencedBy	jpcoar
isRelatedTo	isRelatedTo	Scholix
isReplacedBy	replaces	jpcoar
replaces	isReplacedBy	jpcoar
isRequiredBy	requires	jpcoar
requires	isRequiredBy	jpcoar
isReviewedBy	reviews	DataCite
reviews	isReviewedBy	DataCite
isSupplementedBy	isSupplementTo	jpcoar
isSupplementTo	isSupplementedBy	jpcoar
equivalent	equivalent	CiNii Books
absorption of	absorbed by	CiNii Books
continuation of	continued by	CiNii Books
separated from	continued in part by	CiNii Books
preceded by	succeeded by	CiNii Books

Source	URL
jpcoar	https://github.com/JPCOAR/schema/tree/master/1.0/documents
DataCite	https://schema.datacite.org/
Scholix	http://www.scholix.org/schema
CiNii Books	https://www.jla.or.jp/Portals/0/data/iinkai/mokuroku/ncr2018/ncr2018_c1_201812.pdf

We connected CiNii Research datasets with relational links



Interface improvements

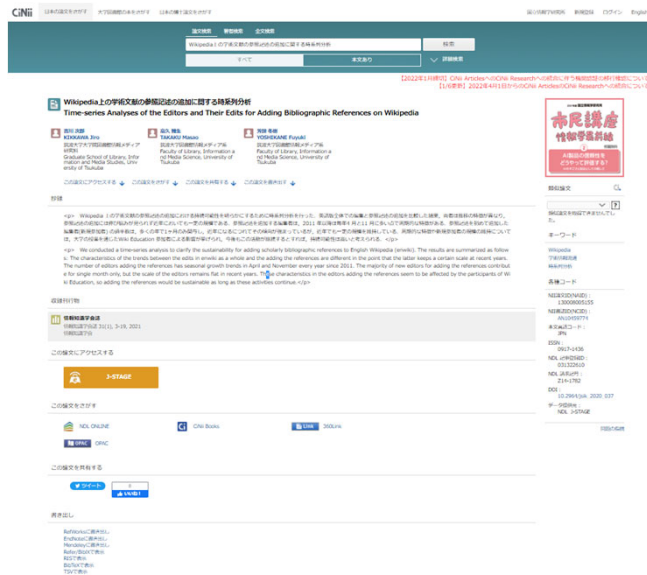
With the data integration and the addition of relation data, we achieved the building a rich and FAIR discovery platform in Japan.

Old interface

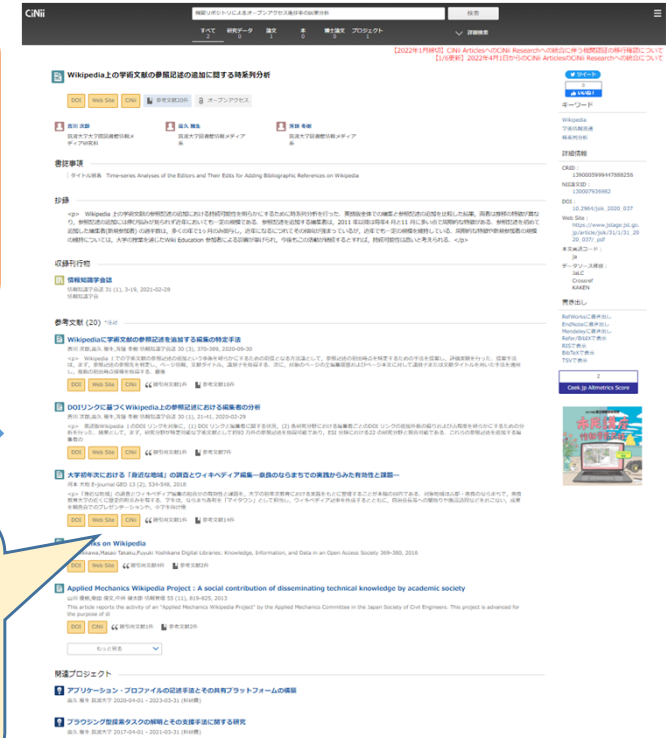


Service Evolution

New interface



* Reference link
* Related datasets
* Related projects
* Related articles
has been added to article page on CiNii Research.



The volume of knowledge available on a single screen became three times bigger

Result

- CiNii Research is the largest national academic discovery platform in Japan
- We executed three steps data processing
 - (Data process 1) Integration using PID
 - (Data process 2) Integration with using text matching and ID mapping
 - (Data process 3) Connecting instances by using citation information and parent-child information.
- This data processing contributed to the richness and the FAIRness of search results of discovery platform.