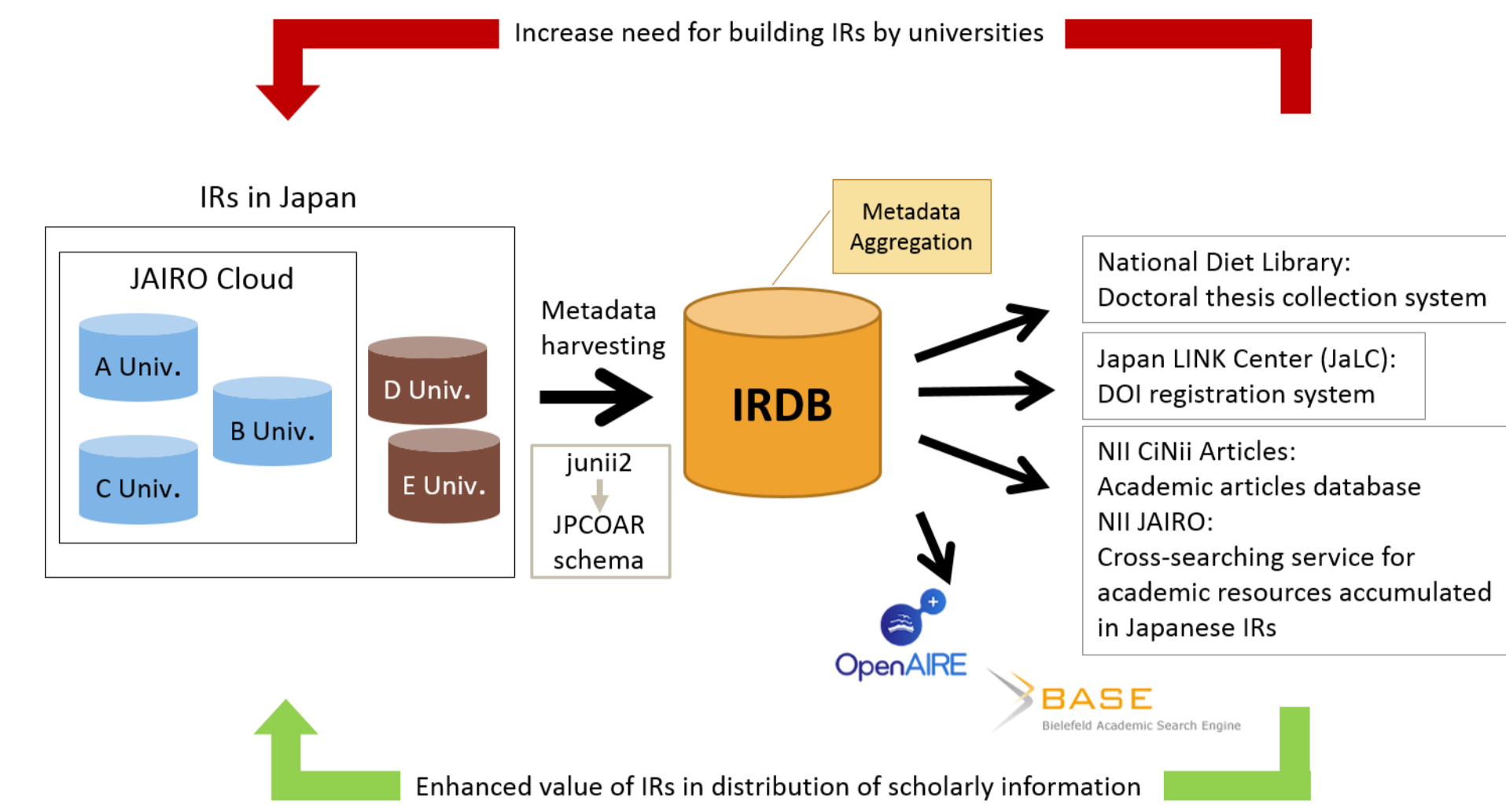




J P C O A R
 Data Rescue Project



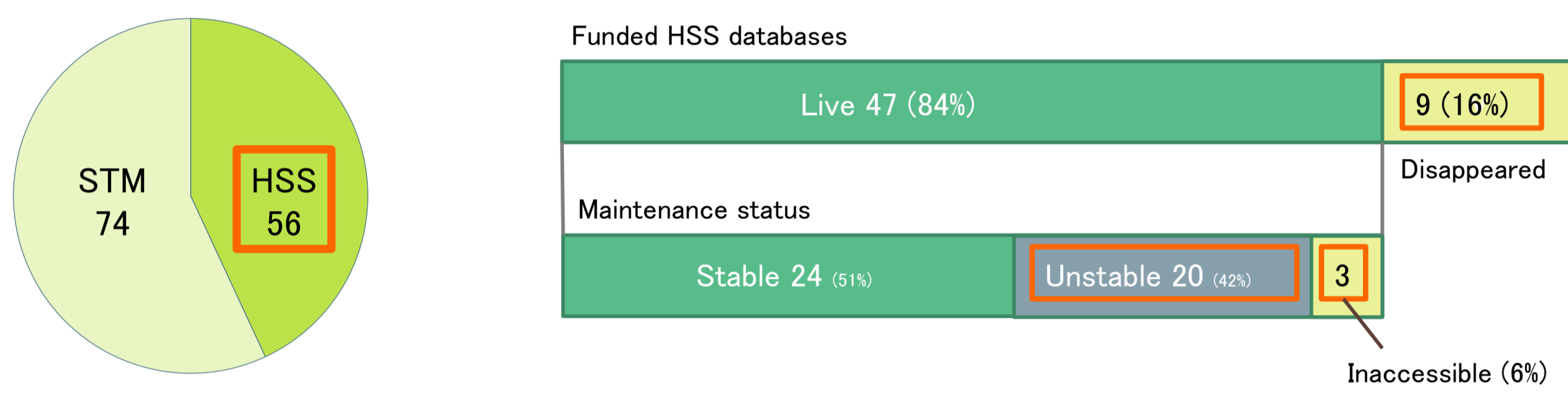
Research Data Task Force



Preliminary Survey

GRANT-IN-AID-FUNDED HSS DATABASES

Database sustainability survey of 130 databases funded by Grant-in-Aid (KAKENHI)* in 2009 - 2001

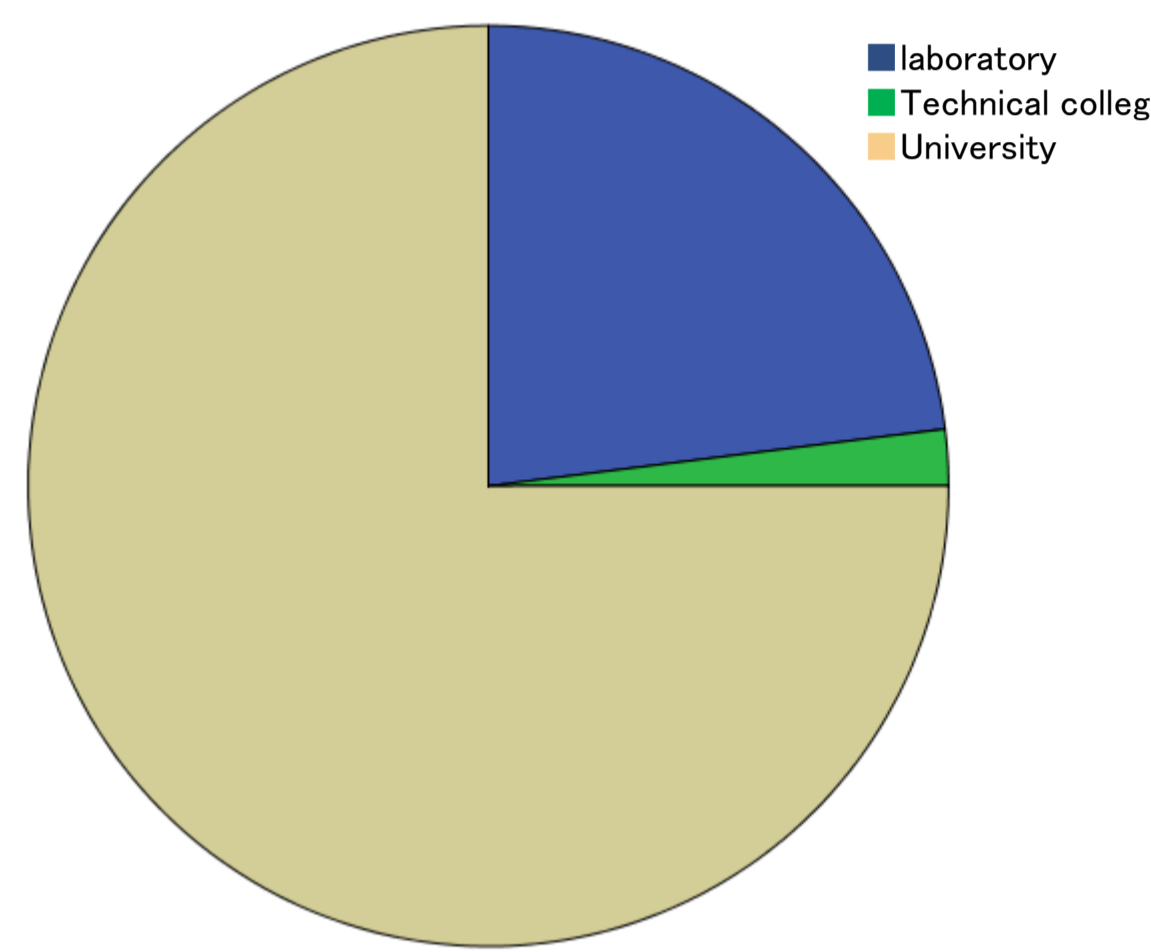


* KAKENHI is the largest public funding programs in Japan. Total amount of funding by KAKENHI as Publication of Scientific Research Results program for databases creation and book publication etc. is 870M JPY (8M USD) in2016

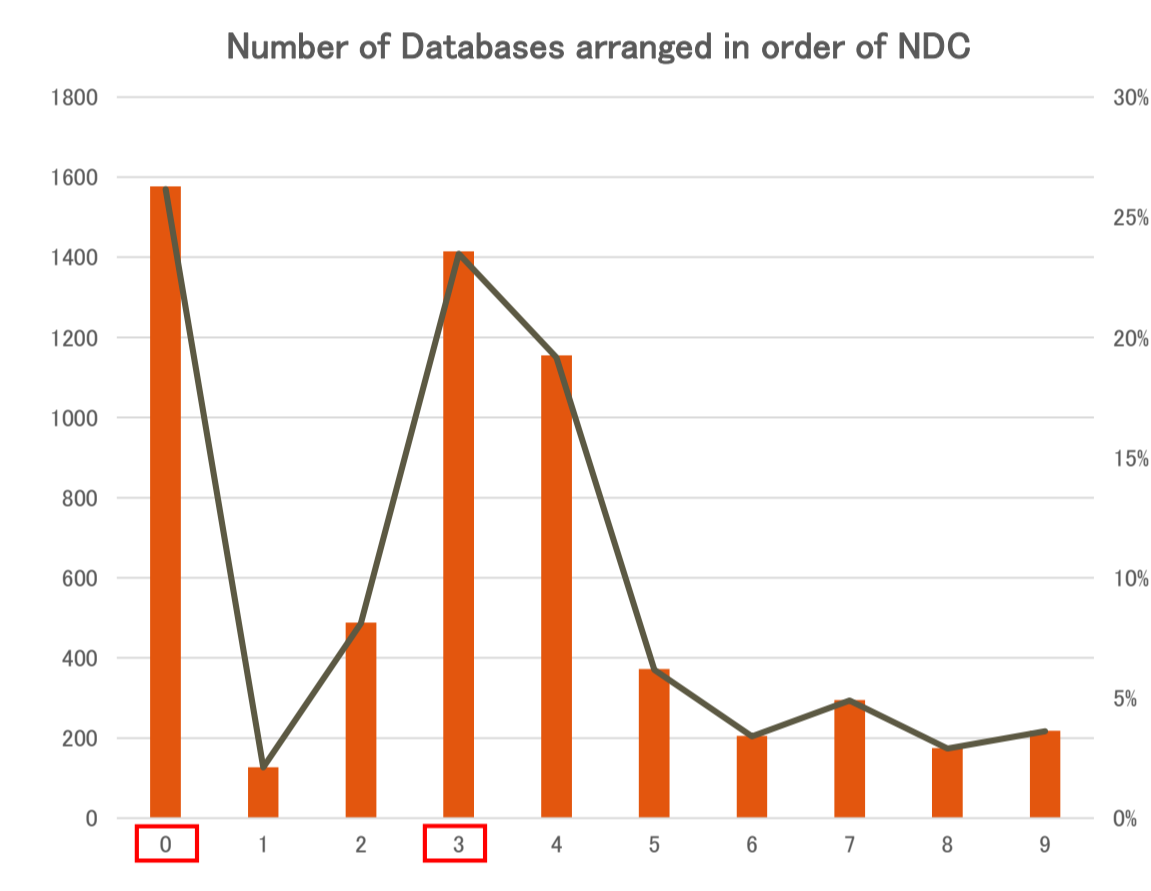
Numerous databases in various fields have been published as the result of publicly funded research in Japan. From FY2009 to FY2011, the Grant-in-Aid for Publication of Scientific Research Results: Databases program, administered by the Japan Society for the Promotion of Science, supported the creation of 130 databases, 74 of which contained research results in the natural sciences and 56 of which contained research results in the humanities and social sciences (HSS).

To examine database sustainability, we conducted a preliminary survey in which we evaluated the sustainability of the 56 HSS Grant-in-Aid-funded databases with respect to their maintenance status and publication methodology. As of April 2015, 47 (84%) of the HSS databases were live, and their maintenance status was as follows: stable, 24 (51%); unstable, 20 (42%); inaccessible (including CD-ROM-based, fee-based, and registration-based databases), 3 (6%). Among the databases with unstable management, we found some that appeared to have not been updated in recent years, and for one of the databases the lead researcher had moved to another institute and the current owner of the database was not specified. Regarding publication methodology, some databases were provided in the legacy FileMaker format or as a single PDF file. Such methodologies do not ensure future accessibility to the database so newer technology should be applied to these databases to prevent loss of the research results they contain.

OTHER DATABASES LISTED IN Dnavi



The National Diet Library Database Navigation Service (Dnavi), formerly managed by the National Diet Library of Japan and currently archived as a static dataset, is a searchable list of databases that includes not only databases published by academic institutions and individual researchers but also public use databases. Approximately 25% (5,333) of the databases in Dnavi are managed by universities, laboratories, and technical colleges. A preliminary survey conducted in 2016 revealed that the majority of databases in Dnavi created after 2,000 were in most cases of discontinuation, and about half of these idle databases contained HSS research results. These databases will likely become inaccessible in the near future resulting in loss of the valuable data they contain. Therefore, to secure access to these databases and prevent data loss, the data should be migrated to a stable infrastructure such as an institutional repository.



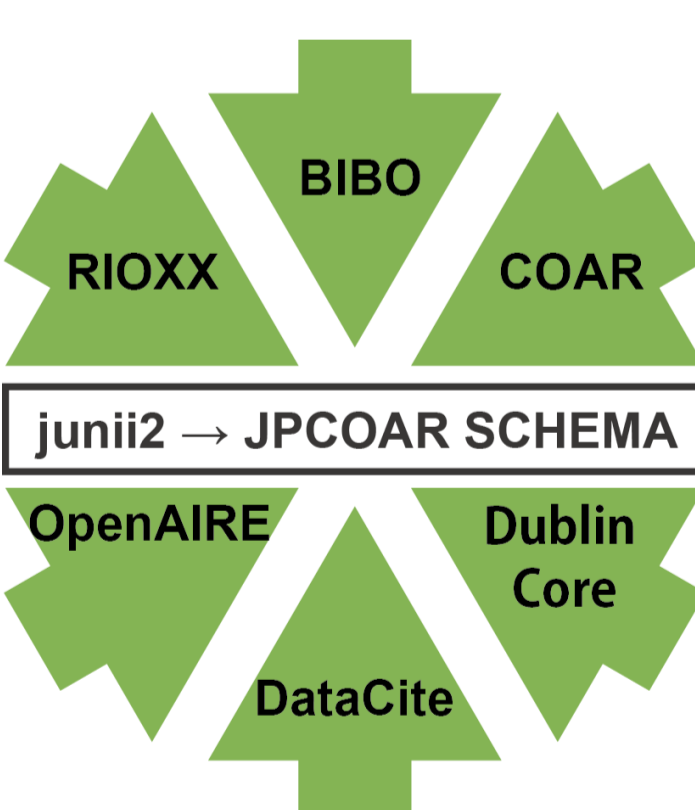
*This figure shows a number of databases managed by universities, laboratories, and technical colleges arranged in order of NDC (Nippon Decimal Classification).

No. 0 (General Works) and No. 3 (Social Science) account for half of those databases.

Challenges

DEVELOPING A WORKFLOW

To migrate unstable databases to institutional repositories, a workflow that is applicable to most databases and streamlines the process of converting and transferring the database metadata to the institutional repository is essential. In addition, a new metadata schema, the JPCOAR metadata schema, built specifically for the Japanese situation was developed recently by the Metadata Task Force of JPCOAR to replace the current standard schema for institutional repositories, juni2. The JPCOAR metadata schema assumes preservation of non-literature materials that were created in accordance to international standards. The JPCOAR metadata schema was developed to ensure that the metadata contained in Japanese institutional repositories is compatible with the Open Science framework and is interoperable internationally.



- * focused on interoperability with six established international metadata standards
- * adopted major identifiers (Crossref Funder ID, ISNI, ORCID, etc.)
- * monitored compliance with open access policies
- * hierarchized priority elements (creator, fundingReference, etc.)
- * enhanced elements related to research data

JuNii2 / JPCOAR schema summary

	juni2	JPCOAR schema
Number of Elements	64	71
Integrated / Abolished	39	—
Add	—	46
Available Number of Attribution	4	13
Available Number of Identifier	12	24
Number of Elements which use Controlled Vocabulary	5	21

*JPCOAR schema is currently being revised. All numbers are in round number as of September 2017.

Mapping examples

Rare Materials		Scrolls		Classical Images	
Elements	Definition	Elements	Definition	Elements	Definition
タイトル	title	書名	Title	資料名	title
タイトルヨミ	transcription	ヨミ	Yomi	資料名ヨミ	titleTranscription
著者	author	作成者	creator		
カテゴリ	category	主題	subject	分類	subject
解説蔵書印	description Ownership Stamp	備考	comments	注記	comment
出版者	Publisher			出版者	publisher
年月日	Date	受入日	Accession Date	作成日	date
		言語	lang	言語	language
カテゴリ	Category	資料タイプ	type	資料種別	type
番号	id	ID	ID	資料番号	identifier
参考資料	reference	関連資料	relation	活字本の巻数 活字本のページ	book vol page

Elements	Definition	Attribute
Title	dc:title	xml:lang
Title	dc:title	xml:lang="ja-Kana"
Creator	jpcoar:creator	jpcoar:creatorName
Subject	jpcoar:subject	xml:lang
Description	datacite:description	xml:lang
Publisher (Distributor)	dc:publisher	xml:lang
Date	datacite:date	
Language	dc:language	
Resource Type	coar:resourceType	
Resource Identifier	datacite:identifier	
Relation	jpcoar:relation	jpcoar:relatedTitle

In FY2016, the Research Data Task Force and Metadata Task Force carried out a trial migration of a Grant-in-Aid-funded database to an institutional repository; after subsequent review of the trial, they proposed the following workflow:

- (1) Select a database in Dnavi that is under unstable management (live or discontinued databases are considered).
- (2) Acquire permission from the database administrator(s) to migrate the contained data to a new database platform.
- (3) Examine the database metadata and select and implement the most appropriate conversion method for registration of the metadata in an institutional repository.
- (4) Map the database's unique metadata schema to the JPCOAR metadata schema.

Database under consideration:

URL: <http://manwe.lib.u-ryukyuu.ac.jp/yanaiharadao/>

Basic information
 Database name: YANAIHARA TADAO collection (Grant-in-Aid-funded databases)
 File format: LDF (LuraDocument), TIFF (for preservation)
 Number of files: 20,473
 File size: about 29MB / file, 560GB / all data
 Number of metadata: over 10,000, each metadata has 10 elements

Special Thanks: University of the Ryukyus Library

FUTURE PLAN

In cooperation with several universities we are now testing our proposed workflow. By following this workflow, we intend to map the metadata from unstable research databases to the JPCOAR metadata schema and register the databases in real-world institutional repositories.

REFERENCES

- Grant-in-Aid for Scientific Research (KAKENHI) <https://www.jsps.go.jp/english/e-grants/index.html>
- National Diet Library Database Navigation Service (Dnavi) <http://doi.org/10.11501/8427554>
- Japan Consortium for Open Access Repository (JPCOAR) <https://jpcoar.repo.nii.ac.jp/>