

# Sample management in JAMSTEC: Cooperation-based management of diverse samples

Takayuki Tomiyama<sup>1</sup>, Tohru Iseto<sup>1</sup>, Yasumi Toyoda<sup>1</sup>, Miki Morioka<sup>1</sup>, and Makoto Ito<sup>1</sup>

<sup>1</sup>JAMSTEC

November 24, 2022

## Abstract

JAMSTEC publicly distributes samples and data obtained by JAMSTEC-AORI-JURCAOS Cooperative Research Cruises [1], as well as those by JAMSTEC private/open-call cruises, through online database (DB) websites, for additional uses with research/educational purposes. These online DBs are, from a long-term perspective, under a process of reorganization and integration: Public-level information systems for rock and sediment core samples were integrated into DARWIN DB [1] in FY2017. As for marine biological samples, whose information has been mainly distributed through “Marine Biological Sample DB [3], a part of information is already shared with DARWIN to utilize its search/map functions. DB integration is in progress for not only public-level but also management-level information system, which are operated by sample holders responsible for storage of physical samples. In FY2018, JSDB (JAMSTEC Sample Management Database) is installed for the management of JAMSTEC rock, sediment core and biological samples, and actual data was transferred into this new system in FY2019. For rock and sediment core samples, information forwarding from management-level system (JSDB) to public-level system (DARWIN) is planned to be established in the near future. JSDB consists of data tables for sample metadata information for onboard sampling activities, sample archive information for status and storage of physical samples, associate data information for various data obtained for samples, literature information for published documents, sample distribution information for service logs for sample requests. User authorization includes group setting, enabling individual sample holders outside Data Management Office to read, create and modify the information in an appropriate manner, although the DB is only accessible from inside JAMSTEC because of security reasons. Associate data information is supposed to include various data types, such as photos/images, text-based descriptions, spread-sheets and URLs etc., and open/close status control and group-setting are available for individual entries. Information for research proposals/projects or other data management issues, which should be associated with multiple samples, can also be managed as associated data information. For the future possibility of involvement with IGSN, sample metadata entries and their vocabulary are reorganized to follow the IGSN Descriptive Metadata [4], as much as possible. Peoples involved in sample and data management tend to spend much effort for dealing with unusual conditions/situations which are not originally expected. Robustness is important for design/operation of information systems, because lack of flexibility and too much complexity of information systems are frequent causes of such kind of problems. [1] <https://www.aori.u-tokyo.ac.jp/coop/> [2] <http://www.godac.jamstec.go.jp/darwin/> [3] <http://www.godac.jamstec.go.jp/bio-sample/> [4] <https://www.geosamples.org/metadata>



# Sample management in JAMSTEC: Cooperation-based management of diverse samples

Takayuki Tomiyama, Tohru Iseto, Yasumi Toyoda, Miki Morioka, Makoto Ito  
and JAMSTEC Sample Management Team

Japan Agency for Marine-Earth Science and Technology (JAMSTEC)  
Data Management Office: dmo@jamstec.go.jp

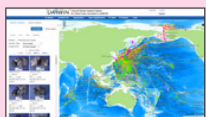


## JAMSTEC Samples & Data

### Public Distribution

JAMSTEC publicly distributes samples and data obtained by research cruises through several websites. Public users can request/download the samples and data, and utilize them for additional uses with research/educational purposes.

### Websites



DARWIN cruise/diver data & samples



J-EDI dive video & photo



Marine biological sample database



BISMAL marine life occurrence



Core sample curation website



And more at NUUNKUI

Variety of websites are operated for distributing information to the public, reflecting the wide variations of information types and the complex workflows in behind. Related information is linked together and create a large network among these websites.

## In-House Sample Management Database

### JAMSTEC Sample Database (JSDB) for Rock, Sediment Core & Biological Sample Management

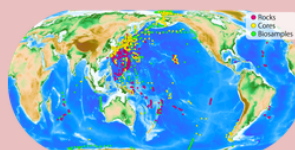
JSDB, an integrated management-level database for JAMSTEC samples was newly installed. These sample types were previously managed using separated information systems.



-System package: I.B.MUSEUM v12, (Waseda Sys. Devel. Co. Ltd.)

-Server: vCPU: 8 x 2.3GHz /MEM: 16GB /HDD: 300GB /CentOS 7.5

### JSDB Sample Information



Statistics of database records (as of 01/07/2020)

Sample Type	Metadata	Physical Samples	Associated Data
Rocks	24413	13563 <sup>1)</sup>	6889
Cores <sup>2)</sup>	1616	963	13343
Biosamples	60277	60276 <sup>1)</sup>	-

<sup>1)</sup>Includes 2811 records of duplicated samples.  
<sup>2)</sup>Includes 273 records of photo-only samples.

List of sample metadata, i.e. the information of onboard sampling activities, is submitted by onboard researchers after each cruise, whereas information of archived physical samples and sample associated data, are provided at different timing by on-shore sample managers, onboard technicians or research groups. Using JSDB, operators can organize the complex relations among these records.

## Development Notes

### Major Features of JSDB

-**Fit to multiple sample types:** Data items for geological samples were re-organized to follow the IGSN metadata schema. Biological data items such as those for taxonomy were incorporated as well.

-**Can handle various data formats:** Electric files with any data format can be uploaded to JSDB, as associated data for samples.

-**Designed for group-based sample management:** Each record is secured by user group write permission, and by flag controls that allow intergroup browsing and on-line publication.

-**Utilize package built-in functions:** The developer package accommodates convenient user interfaces, such as input assistance, complex data search and customizable viewers.

### Future Plans and Problems

-**Data forwarding to public distribution systems:** Implementation of data synchronization module for DARWIN is on-going. The module re-maps data items to compensate the differences in metadata schema.

-**Further commitment with IGSN:** Currently, limited samples have IGSNs as URL-type associated data. Major involvement with IGSN will need detailed considerations for information workflows, including ID allocation, information sharing with researchers and IGSN administration office.

-**In-house collaboration:** Physical samples are independently managed in multiple departments and research groups. In-house information sharing is widely operated for biological samples. Similar collaboration may help effective management of rocks and cores.

## Sample management in JAMSTEC: Cooperation-based management of diverse samples

Takayuki Tomiyama, Tohru Iseto, Yasumi Toyoda, Miki Morioka, Makoto Ito, and JAMSTEC  
Sample Management Team  
Japan Agency for Marine-Earth Science and Technology (JAMSTEC)  
Data Management Office: [dmo@jamstec.go.jp](mailto:dmo@jamstec.go.jp)

### JAMSTEC Samples & Data

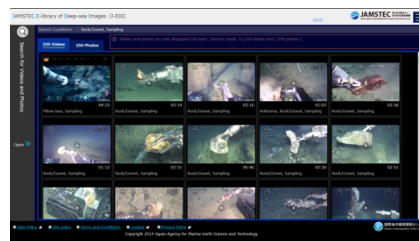
#### Public Distribution

JAMSTEC publicly distributes samples and data obtained by research cruises through several websites. Public users can request/download the samples and data, and utilize them for additional uses with research/educational purposes.

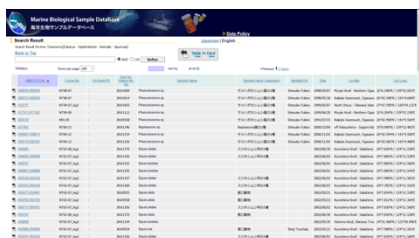
#### Websites



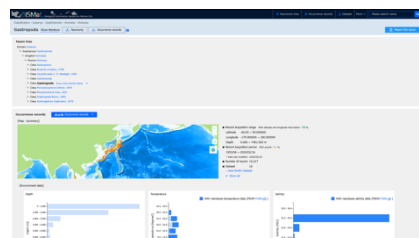
DARWIN cruise/dive data & samples  
<http://www.godac.jamstec.go.jp/darwin/e>



J-EDI dive video & photo  
<https://www.godac.jamstec.go.jp/jedi/e/>



Marine biological sample database  
[http://www.godac.jamstec.go.jp/bio-sample/index\\_e.html](http://www.godac.jamstec.go.jp/bio-sample/index_e.html)



BISMAL marine life occurrence  
<https://www.godac.jamstec.go.jp/bismal/e/>



Core sample curation website  
[http://www.jamstec.go.jp/kochi/jc\\_curation/e/](http://www.jamstec.go.jp/kochi/jc_curation/e/)



And more at NUUNKUI  
<http://www.godac.jamstec.go.jp/jmedia/portal/e/>

Variety of websites are operated for distributing information to the public, reflecting the wide variations of information types and the complex workflows in behind. Related information is linked together and create a large network among these websites.

## In-House Sample Management Database

### JAMSTEC Sample Database (JSDB) for Rock, Sediment Core & Biological Sample Management

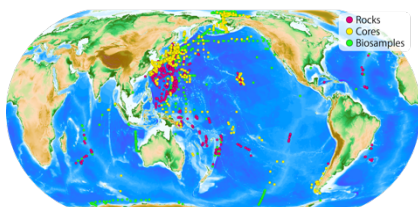
JSDB, an integrated management-level database for JAMSTEC samples was newly installed. These sample types were previously managed using separated information systems.



Screenshot of JSDB

- System package: I.B.MUSEUM v12, (Waseda Sys. Devel. Co. Ltd.)
- Server: vCPU: 8 x 2.3GHz /MEM: 16GB /HDD: 300GB /CentOS 7.5

### JSDB Sample Information



Global map of JAMSTEC sampling activities

Statistics of database records (as of 01/07/2020)

Sample Type	Metadata	Physical Samples	Associated Data
Rocks	24413	13563 <sup>*1</sup>	6889
Cores <sup>*2</sup>	1616	963	13343
Biosamples	60277	60276 <sup>*3</sup>	-

<sup>\*1</sup>Includes 2811 records of duplicated samples.

<sup>\*2</sup>Numbers refer to coring events or drilled holes.

<sup>\*3</sup>Includes 273 records of photo-only samples.

Table for JSDB record statics

List of sample metadata, i.e. the information of onboard sampling activities, is submitted by onboard researchers after each cruise, whereas information of archived physical samples and sample associated data, are provided at different timing by on-shore sample managers, onboard technicians or research groups. Using JSDB, operators can organize the complex relations among these records.

## Development Notes

### Major Features of JSDB

- Fit to multiple sample types: Data items for geological samples were re-organized to follow the IGSN metadata schema. Biological data items such as those for taxonomy were incorporated as well.
- Can handle various data formats: Electric files with any data format can be uploaded to JSDB,

as associated data for samples.

- Designed for group-based sample management: Each record is secured by user group write permission, and by flag controls that allow intergroup browsing and on-line publication.
- Utilize package built-in functions: The developer package accommodates convenient user interfaces, such as input assistance, complex data search and customizable viewers.

### **Future Plans and Problems**

- Data forwarding to public distribution systems: Implementation of data synchronization module for DARWIN is on-going. The module re-maps data items to compensate the differences in metadata schema.
- Further commitment with IGSN: Currently, limited samples have IGSNs as URL-type associated data. Major involvement with IGSN will need detailed considerations for information workflows, including ID allocation, information sharing with researchers and IGSN administration office.
- In-house collaboration: Physical samples are independently managed in multiple departments and research groups. In-house information sharing is widely operated for biological samples. Similar collaboration may help effective management of rocks and cores.