



# Data Management Plan of DDBJ Datasets using Shiny-SurrealDB: a study case

Andrea Ghelfi and Takatomo Fujisawa

第27回オープンバイオ研究会

2023-03-11



# Overview

INTRODUCTION  
3

---

DATA MANAGEMENT  
GUIDELINES  
4

---

DATABASE SECURITY: ISSUES  
TO CONSIDER  
5

---

STUDY CASE:  
SURREAL VIEWER  
6

---

SURREALDB  
DATA INTEGRATION  
10

---

SUMMARY  
11

# Introduction

At DDBJ we deal with a number of databases which occasionally requires new platforms for an easy access, store and integrate a particular dataset.

Having in mind that lose data is a historical problem in scientific research, and even when data is not lost, in some cases precious research time is lost when trying to find or understand what the it means.

Therefore, we understand that implementation of a data management plan should be aimed, so less time finding, understanding, reusing and/or sharing data would be spent by researchers. As well as, reliable procedures should be implemented in order to deal with private data.

# Data Management Guidelines

|               |                           |   |
|---------------|---------------------------|---|
| Documentation | How To                    | Record the most important information, in README files, codebooks, notebooks, such as GitHub.   |
| Consistency   | Where                     | Organize folders, subfolders and filenames in a logical structure, such as by project and/or date.                                    |
| Version       | When                      | Keeping distinct copies, for example keep original files and the clean final version, in a classification system such as Git.         |
| Security      | Who can/<br>Who shouldn't | (a) Establish security controls for accession of confidential data. (b) Security breach, such as SQL injection attacks, Malware, etc. |
| Constraints   | Which                     | Identify best tools based on evaluation of accessibility, speed, vulnerabilities and limitations.                                     |
| Back-up       | Why                       | Automated and with regular frequency, daily or weekly. Also, perform periodical confirmation that backups are working properly.       |

# Database security: Issues to consider

- Encrypt confidential data, such as masking and tokenization should be considered.
- Access to database should be restricted to the minimum level.
- Web applications that access the database should comply with best practices security guidelines.
- Monitoring log activity and automated detection of suspicious activities.

The background features a light gray base with large, organic, overlapping shapes in muted olive green and dusty rose. In the top left corner, there is a stylized, light gray illustration of a pine branch with needle-like leaves. Two thin, white, curved lines sweep across the bottom right portion of the image.

# Study Case: Surreal Viewer

Surreal Viewer

View

About

Query type

Project

Query ID

PRJDB5665

Download Data

Download

Trad

Show 10 entries

Search:

| Accession | Project   | BioSample    | db       | Taxon   | Status | Count |
|-----------|-----------|--------------|----------|---------|--------|-------|
| AP018166  | PRJDB5665 | SAMD00079794 | g-actual | 272123  | public | 1     |
| AP018167  | PRJDB5665 | SAMD00079794 | g-actual | 272123  | public | 1     |
| AP018168  | PRJDB5665 | SAMD00079794 | g-actual | 272123  | public | 1     |
| AP018169  | PRJDB5665 | SAMD00079794 | g-actual | 272123  | public | 1     |
| AP018170  | PRJDB5665 | SAMD00079794 | g-actual | 272123  | public | 1     |
| AP018171  | PRJDB5665 | SAMD00079794 | g-actual | 272123  | public | 1     |
| AP018172  | PRJDB5665 | SAMD00079805 | g-actual | 1954171 | public | 1     |
| AP018173  | PRJDB5665 | SAMD00079805 | g-actual | 1954171 | public | 1     |
| AP018174  | PRJDB5665 | SAMD00079795 | g-actual | 1085406 | public | 1     |
| AP018175  | PRJDB5665 | SAMD00079795 | g-actual | 1085406 | public | 1     |

Showing 1 to 10 of 172 entries

Previous12345...18Next

Trace

Show 10 entries

Search:

| Project   | project_type | BioSample    | Taxon   | DRR       | bp_submitter | bs_submitter | bp_status | bs_status | bp_locus_tag | bs_locus_tag |
|-----------|--------------|--------------|---------|-----------|--------------|--------------|-----------|-----------|--------------|--------------|
| PRJDB5665 | primary      | SAMD00079794 | 272123  | DRR315817 | tfuji        | tfuji        | public    | public    | NIES         | NIES19       |
| PRJDB5665 | primary      | SAMD00079794 | 272123  | DRR315817 | tfuji        | tfuji        | public    | public    | NIES         | NIES19       |
| PRJDB5665 | primary      | SAMD00079794 | 272123  | DRR315816 | tfuji        | tfuji        | public    | public    | NIES         | NIES19       |
| PRJDB5665 | primary      | SAMD00079794 | 272123  | DRR315816 | tfuji        | tfuji        | public    | public    | NIES         | NIES19       |
| PRJDB5665 | primary      | SAMD00079795 | 1085406 | DRR315818 | tfuji        | tfuji        | public    | public    | NIES         | NIES21       |

Surreal Viewer

View

About

Query type

BioSample

Query ID

SAMD00079800

Download Data

Download

Trad

Show 10 entries

Search:

| Accession | Project   | BioSample    | db       | Taxon   | Status | Count |
|-----------|-----------|--------------|----------|---------|--------|-------|
| AP018314  | PRJDB5665 | SAMD00079800 | g-actual | 1973480 | public | 1     |
| AP018315  | PRJDB5665 | SAMD00079800 | g-actual | 1973480 | public | 1     |

Showing 1 to 2 of 2 entries

Previous1Next

Trace

Show 10 entries

Search:

| Project   | project_type | BioSample    | Taxon   | DRR       | bp_submitter | bs_submitter | bp_status | bs_status | bp_locus_tag | bs_locus_tag |
|-----------|--------------|--------------|---------|-----------|--------------|--------------|-----------|-----------|--------------|--------------|
| PRJDB5665 | primary      | SAMD00079800 | 1973480 | DRR315871 | tfuji        | tfuji        | public    | public    | NIES         | NIES73       |
| PRJDB5665 | primary      | SAMD00079800 | 1973480 | DRR315873 | tfuji        | tfuji        | public    | public    | NIES         | NIES73       |
| PRJDB5665 | primary      | SAMD00079800 | 1973480 | DRR315873 | tfuji        | tfuji        | public    | public    | NIES         | NIES73       |
| PRJDB5665 | primary      | SAMD00079800 | 1973480 | DRR315872 | tfuji        | tfuji        | public    | public    | NIES         | NIES73       |
| PRJDB5665 | primary      | SAMD00079800 | 1973480 | DRR315872 | tfuji        | tfuji        | public    | public    | NIES         | NIES73       |
| PRJDB5665 | primary      | SAMD00079800 | 1973480 | DRR315871 | tfuji        | tfuji        | public    | public    | NIES         | NIES73       |

Showing 1 to 6 of 6 entries

Previous1Next



Surreal Viewer

localhost:10000/p/bb0dc5a7/

120%

Surreal Viewer

View

About

Trad

Show 10 entries

Search:

| Project   | BioSample    | db       | Taxon   | Status | Count |
|-----------|--------------|----------|---------|--------|-------|
| PRJDB5665 | SAMD00079811 | g-actual | 1137095 | public | 1     |
| PRJDB5665 | SAMD00079811 | g-actual | 1137095 | public | 1     |
| PRJDB5665 | SAMD00079811 | g-actual | 1137095 | public | 1     |
| PRJDB5665 | SAMD00079811 | g-actual | 1137095 | public | 1     |
| PRJDB5665 | SAMD00079811 | g-actual | 1137095 | public | 1     |
| PRJDB5665 | SAMD00079811 | g-actual | 1137095 | public | 1     |
| PRJDB5665 | SAMD00079811 | g-actual | 1137095 | public | 1     |
| PRJDB5665 | SAMD00079811 | g-actual | 1137095 | public | 1     |

8 entries

Previous1Next

Query type

Project

Project

BioSample

Accession

Taxon

DRR

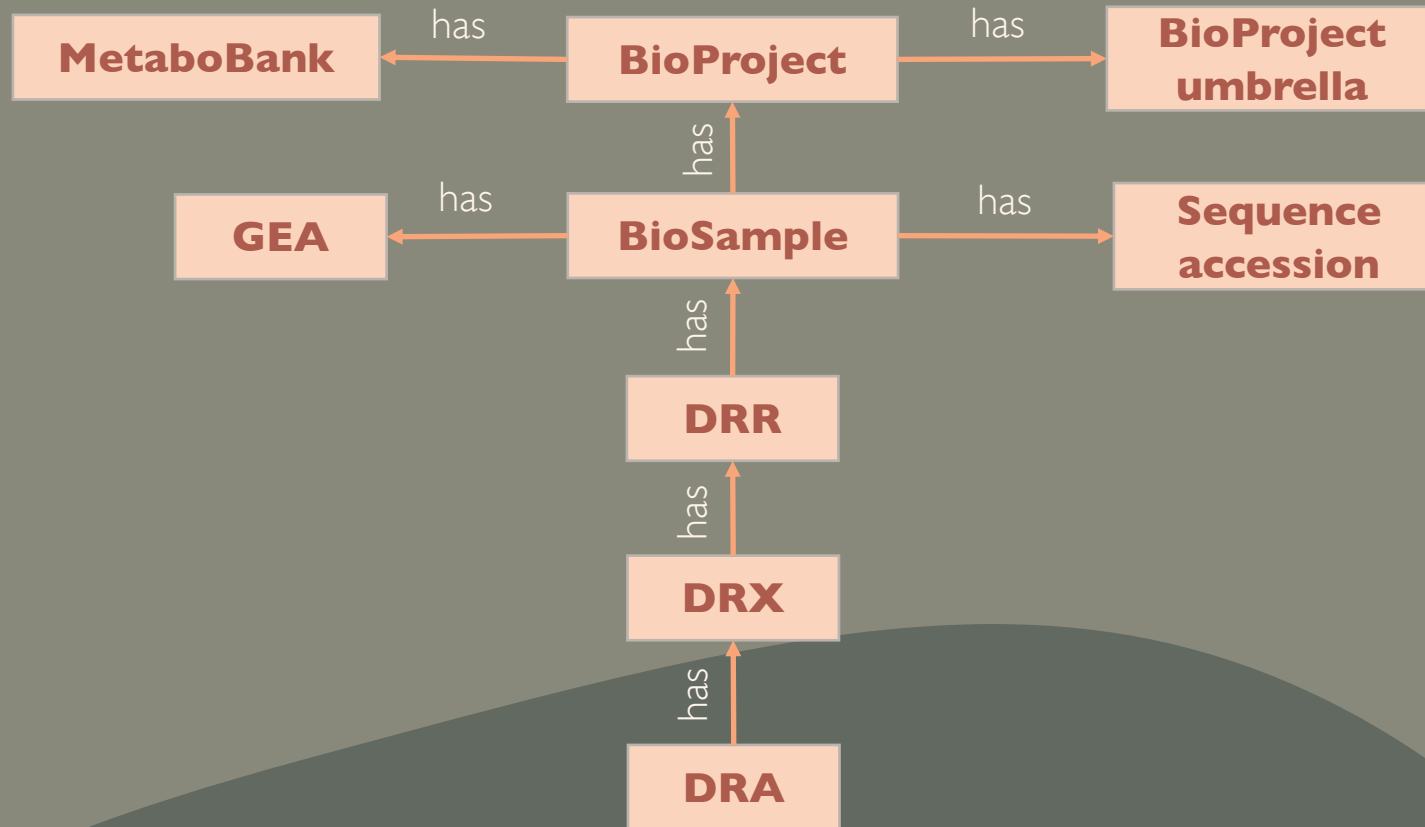
Download Data

Download

Search:

| Project   | project_type | BioSample    | Taxon   | DRR       | bp_submitter | bs_submitter | bp_status | bs_status | bp_locus_tag | bs_locus_tag |
|-----------|--------------|--------------|---------|-----------|--------------|--------------|-----------|-----------|--------------|--------------|
| PRJDB5665 | primary      | SAMD00079811 | 1137095 | DRR315867 | tfuji        | tfuji        | public    | public    | NIES         | SAMD00079811 |
| PRJDB5665 | primary      | SAMD00079811 | 1137095 | DRR315868 | tfuji        | tfuji        | public    | public    | NIES         | SAMD00079811 |
| PRJDB5665 | primary      | SAMD00079811 | 1137095 | DRR315868 | tfuji        | tfuji        | public    | public    | NIES         | SAMD00079811 |

# SurrealDB Data Integration



# Summary

In order to have a efficient data management there are some important steps to follow, such as:

- Good documentation of the procedures used to manage and store data;
- Establish rules for directory, sub-directories and filenames, which should be easily understandable;
- Versioning all updates in a structured manner such as used in GitHub;
- Secure confidential data as well as being aware of security breaches;
- Identify strengths and weakness of the tools that have being used in the system;
- Automate a periodic backup to secure data.

# References

1. Briney KA, Coates H, Goben A (2020) Foundational Practices of Research Data Management. Research Ideas and Outcomes 6: e56508.
2. Database security - <https://www.ibm.com/topics/database-security>.
3. Database management - <https://www.smartsheet.com/database-management>.



# ありがとうございます

Andrea Ghelfi

[andreaghelfi@nig.ac.jp](mailto:andreaghelfi@nig.ac.jp)