SWISSUbase Data Versioning Policy

Table of Contents

1. Introduction

Research data should be easily and reliably identified and cited. Additionally, SWISSUbase users must trust the origin and reliability of data objects in the catalogue. Therefore, version control is essential to track changes and establish the state of a resource, which is especially critical for digital objects that are frequently revised and updated.

SWISSUbase differentiates between studies and datasets. Studies describe the overall scope of the project and can have one or multiple datasets. Datasets detail the specific data used or recorded during a project and contain actual data files. While both studies and datasets are versioned within SWISSUbase, only datasets receive DOIs.

This policy applies to all datasets and studies recorded in SWISSUbase as of June 18, 2024.[1] It outlines the principles and procedures for managing versioning to ensure consistent tracking and clear identification of data revisions.

2. DOI – Digital Object Identifier

Persistent Identifiers (PIDs) such as Digital Object Identifiers (DOIs)[2] ensure that research data published in SWISSUbase are easily citable in scientific journals and shareable with other researchers. Additionally, using PIDs facilitates making the data better findable,

---

[1] As of June 18, 2024, the new versioning system has updated all SWISSUbase dataset version numbers to two digits, affecting both new and previously published datasets. Note: Citations for datasets published before this date retain their original three-digit version numbers, now called "legacy version numbers".

[2] DataCite: DOI Basics. https://support.datacite.org/docs/doi-basics [12.06.2024]

accessible, interoperable, and reusable, adhering to FAIR principles[3]. In SWISSUbase, only datasets receive DOIs.

Assignment of DOIs to datasets

- **DOI for each dataset version:** A unique DOI is assigned to each newly created dataset and to each major update that significantly alters the dataset's content or structure, ensuring accurate tracking and citation.

- **Same DOI for minor changes:** Minor changes to a dataset do not generate a new DOI, reducing the number of DOIs per dataset and simplifying citation.

- **Canonical DOI for datasets:** A canonical DOI is assigned for the entire dataset, which allows users to cite the dataset as a whole regardless of version changes. This DOI points to the most current version of the dataset but provides access to all versions through the version history.

All DOIs, including those for significant updates and the canonical DOI, are visible in the dataset's version history.

3. Versioning Principles

SWISSUbase bases its versioning principles on the frameworks provided by Semantic Versioning[4], DataCite's versioning recommendations[5], and the Research Data Alliance (RDA) Data Versioning Working Group[6].
These frameworks ensure that specific versions of data used to support research results are unambiguously referenced and retrievable.

The Versioning framework uses Major, Minor and Cosmetic changes to track and document updates made to a study or dataset on SWISSUbase.

- **Major Changes:** comprehensive alteration to the study's scope or substantial modifications that affect the dataset's content or analytical value.
- **Minor Changes:** Updates or corrections that do not fundamentally alter the study's scope or dataset's content.
- **Cosmetic Changes:** Minor edits, such as spelling or formatting corrections and minor metadata adjustments, that do not substantially alter the metadata.

4. Version Number Structure

When publishing research data, SWISSUbase uses a two-digit versioning scheme according to the pattern "major.minor" (e.g., Version 1.3).

First publications of a dataset on SWISSUbase bear the version number "1.0".

---

[3] Go FAIR: FAIR Principles. https://www.go-fair.org/fair-principles/ [12.06.2024]
[4] Preston-Werner, Tom: Semantic Versioning 2.0.0. https://semver.org/ [12.06.2024]
[5] DataCite: Versioning. https://support.datacite.org/docs/versioning [12.06.2024]
[6] Klump, J., Wyborn, L., Wu, M., Martin, J., Downs, R.R. and Asmi, A., 2021. Versioning Data Is About More than Revisions: A Conceptual Framework and Proposed Principles. Data Science Journal, 20(1), p.12. DOI: http://doi.org/10.5334/dsj-2021-012 [12.06.2024]

Major changes are considered consequential, so the primary version number is incremented: 1.x → 2.0. If the "major" version number is changed, the "minor" version number is numbered again starting from "0", moving upwards (e.g., Version change from 2.14 to 3.0).

Minor changes are considered small-scale, so only the secondary version number is incremented: x.1 → x.2. The second digit can be incremented as high as necessary, without a limit (x.10 … x.30… x.204).

Cosmetic changes are spelling and formatting corrections and other minor alterations to the catalogue metadata entries, which <u>do not substantially alter</u> metadata or documentation. Therefore, they have no influence on the versioning number (e.g., 1.1. remains 1.1).

4.1. Types of Changes

4.1.1. Major changes

Dataset level

Major changes alter the informative value of the research data and lead to the republication of the dataset as a new version with a new DOI.

Examples include major corrections, adjustments or additions to a research dataset, such as deleting, inserting or correcting data and documentation files, which might lead to changes in the analysis or interpretations, additional analyses or updated figures.

Major changes result in the following updates in the system workflow:
- New AIP and DIP are created.
- New major version is created (e.g., 1.x to 2.0).
- New DOI is issued.
- New usage license needs to be selected

Study level

A Major Change involves a comprehensive alteration to the study's scope, such as entirely new project objectives or methodologies. This option, while rare, allows a data depositor the option to update the project description rather than creating a new study from scratch.

Major changes result in a new study version (e.g., 1.0 to 2.0).

4.1.2. Minor changes

Dataset level

Minor changes do not alter the informative value of research data. However, they involve relevant alterations to the catalogue metadata and documentation.

These changes mainly focus on any alterations made to documentation and other additional files independent of the research data, including relevant changes to the catalogue metadata entries.

Minor changes result in the following updates in the system workflow:
- New AIP and DIP are created.
- New minor version is created (e.g., x.3 to x.4).
- <u>No</u> new DOI is issued.

Study level

Minor changes on study level include relevant updates or changes to metadata, such as titles, authors or abstract. It can also include adding additional information about the general description, funding, bibliographical references, or domain specific metadata, as well as additional translations.

Minor changes result in a new study version (e.g., 1.1 to 1.2).

4.1.3. Cosmetic Changes

Dataset and study level

Cosmetic changes involve spelling and formatting corrections and minor alterations to the catalogue metadata entries that do not substantially alter metadata. These changes do not lead to a new version or a new DOI, and the AIP and DIP packages are not updated.

Cosmetic changes can only be implemented by curators. However, researchers can contact data curators to request cosmetic changes.

While the type and manner of these changes do not need to be described in a specific version notes field, a record of the updated cosmetic changes is stored in the application.

4.2. Version control procedures

Both data depositors (researchers) and data curators can initiate updates to published datasets, classified as either Major or Minor changes based on their impact. Cosmetic changes, such as spelling corrections or formatting adjustments, are reserved to data curators as they involve direct intervention in the application.

All proposed changes by data depositors undergo a thorough review by data curators to ensure they adhere to SWISSUbase standards and maintain data integrity. This review is essential to validate the accuracy and necessity of the changes.

Distinguishing between Major and Minor changes can be challenging when updates do not clearly fall into predefined categories. The initial classification made by data depositors is reviewed by curators who have the discretion to reclassify the change if necessary. This process ensures that the classification of changes accurately reflects their impact on the dataset or study.

Changes must be documented by both data depositors and curators in the "Version notes" field of the SWISSUbase metadata schemas for both studies and datasets. This documentation should clearly describe the nature and impact of the changes.

4.3. Version History Management

New or altered versions of a dataset shall be linked to older dataset versions, and the data version history is discoverable in the SWISSUbase online catalogue. The version history of each dataset and study is meticulously documented within SWISSUbase, ensuring that each iteration is linked and traceable.

In the rare case a dataset is deleted, the dataset landing page will be saved with information about the reason for its deletion and a link to the latest published dataset if available (tombstone page).

4.4. Citation

Each published dataset is provided with a data citation from SWISSUbase. This includes a DOI, a reference to the institution who curated the data, as well as the dataset version number to reference each dataset specifically.

The citation used in SWISSUbase is based on the APA-citation style:

Author, A. (Year). Title of the data set (Version number) [Data set]. Publisher Name. DOI

Example:
Dupont, J., Mustermann, M., & Doe, J. (2024). Adaptation of version control procedures to research data (Version 1.0) [Data set]. UNIL data service. https://doi.org/99.00000/SUB-Mock1

With the implementation of our new versioning system on June 18, 2024, all datasets within SWISSUbase received a two digit version number, reflecting a necessary system-wide change. This applies to both current and previously published datasets. It is important to note, however, that the citation format for datasets published before this date remains unchanged, and these datasets will continue to be cited using their original three-digit legacy version numbers.

4.5. Compliance and Monitoring

To ensure adherence to the SWISSUbase Data Versioning Policy, data curators oversee the application of versioning procedures and ensure accurate documentation. Regular reviews and audits are conducted to maintain data integrity. The policy is reviewed and updated periodically to stay aligned with best practices and evolving research data management standards.