MEDIN DAC Versioning Guidance

1 Definitions and Terms

Term	Definition
Revision[1]	• A new instance of a dataset that is produced in the course of data production or data management that is different from its precursor is called a “revision”. • A dataset revision should be identified.
Provenance[1]	• The definition of revisions and releases signifies that a dataset has been derived from a precursor and is part of the description of its lineage, or provenance. • Provenance can be more complex than following a linear path. Information accompanying a dataset release should therefore contain information on the provenance of a dataset.
Release[1]	• The release of a new version of a dataset should be accompanied by a description of the nature and the significance of the change. • The significance of this change will depend on the intended use of the data by its designated user community. • Each new release of a data product should have a new identifier.
Version	• A new occurrence of an object such as a research artefact (data, physical sample, software etc)
Granularity[1]	• Data may be aggregated and combined into collections or timeseries. • The collection should be identified and versioned, as should be each of its constituent datasets. • Entire time series should be identified, as should be time-stamped revisions.
Manifestation[1]	• The same dataset may be expressed in different file formats or character encodings without differences in content. While these datasets will have different checksums, the work expressed in these datasets does not differ, they are manifestations of the same work. • Manifestations of the same work should be individually identified and related to their parent work.
Citation[1]	• Include information of the Release in the citation. DataCite recommends to use semantic versioning, issue a new identifier with major releases, use the “alternate identifier” and “related identifier” elements to identify releases and how they relate to other datasets, e.g. whether it was derived from a precursor. • Updating the metadata does not create a new version, it only changes the catalogue entry.

2 Best Practice

2.1 DOIs

2.1.1 DataCite Schema Guidance

DataCite Schema has its own defined section to determine version (https://datacite-metadata-schema.readthedocs.io/_/downloads/en/4.6/pdf/ p. 41). DataCite suggested practice is to have the following format for versions: ***major_version.minor_version***. A new identifier is to be registered for a major version change. Any determination on whether a change is major vs minor change is with the data stewards involved.

2.1.2 DataCite Versioning Guidance

[1] Klump, J., Wyborn, L., Wu, M., Downs, R., Asmi, A., Ryder, G., & Martin, J. (2020a). Principles and best practices in data versioning for all datasets big and small (1.1). Zenodo. https://doi.org/10.15497/RDA00042

https://support.datacite.org/docs/versioning

When content underlying a DOI is updated, DataCite recommend updating the DOI metadata and, for major changes, assigning a new DOI.

- For minor content changes, the same DOI may be used with updated metadata. A new DOI is not required.
- For major content changes, DataCite recommend assigning a new DOI and linking the new DOI to the previous DOI with related identifiers.

2.2 Dataset versioning

The Global Biodiversity Information Facility (GBIF) Integrated Publishing Toolkit uses the DataCite recommendations for versioning (https://ipt.gbif.org/manual/en/ipt/latest/versioning), specifying that the version number also uses the syntax major_version.minor_version.

2.3 Metadata versioning

The RDA guidance on metadata versioning specifies that if the metadata is changed, due to a correction of the metadata or addition of metadata, if these changes do not affect the dataset, this would not constitute a new dataset version (Klump et al. 2020a).

3 Useful links to information on Versioning

Klump, J., Wyborn, L., Downs, R., Asmi, A., Wu, M., Ryder, G., Martin, J. (2020a). Principles and best practices in data versioning for all data sets big and small. Version 1.1. Research Data Alliance. DOI: 10.15497/RDA00042.

Klump, J., Wyborn, L., Downs, R., Asmi, A., Wu, M., Ryder, G., & Martin, J. (2020b). Compilation of Data Versioning Use cases from the RDA Data Versioning Working Group. Version 1.1. Research Data Alliance. DOI: 10.15497/RDA00041

ESIP Data Preservation and Stewardship Committee. 2019. Data Citation Guidelines for Earth Science Data. Ver. 2. Earth Science Information Partners. https://doi.org/10.6084/m9.figshare.8441816

DataCite Versioning, https://support.datacite.org/docs/versioning. Accessed 20th August 2025.

Global Biodiversity Information Facility (GBIF) Data versioning Policy, https://ipt.gbif.org/manual/en/ipt/latest/versioning. Accessed 20th August 2025.

4 Caveats for the DACs

Emailed GBIF with concerns around their citation creation and use: https://github.com/gbif/ipt/issues/2778

5 Involvement in RDA

If you are interested in being involved in any of the Research Data Alliance (RDA) groups looking at versioning, the following links may be useful:

- Data Citation Working Group
- Complex Citation Implementation Interest Group

- Data Versioning Interest Group

6 Forward look at Versioning

6.1 Complex citations

Complex Citation Objects are a mechanism that can utilise DOIs and PIDs to forma a base level that allows users to interact, adapt, or amend data to create a product or dataset from existing research artefacts. As complex citation objects are metadata signposts for data used, complex citation objects could prevent millions of subtle variations, mutations or divergent datasets from being created and stored as new versions due to requirements on transparency and traceability.

7 Review of Versioning Guidelines for the DACs

This document will be reviewed every two years, with the DACs, to ensure it stays up to date and relevant