

Reporting Missing Values

The International Nucleotide Database Collaboration (INSDC) have a standardised missing/null value reporting language to be used where a value of an expected format for sample metadata reporting **can not** be provided.

The controlled vocabulary takes into account different type of constraints. Submitters are **strongly** encouraged to always provide true values. However, if missing/null value reporting is required, submitters are asked to use a term with the finest granularity for their situation. See the table below for accepted missing value reporting terms.

Note

If your sample metadata does not provide enough context for your data to be easily interpreted, you may be requested to update your samples, so it is important to ensure you provide any metadata available and only use missing value terms when necessary.

INSDC Missing Value Reporting Terms

INSDC term (top level)	INSDC term (lower level)	Definition	INSDC term (reporting level)	Definition
not applicable		information is inappropriate to report, can indicate that the standard itself fails to model or represent the information appropriately	control sample	Information is not applicable as the sample represents a negative control sample collected in a lab
			sample group	Information is not applicable as the sample represents a group of samples that do not have a single origin. E.g. for co-assembly or transcriptome assembly.
missing	not collected	information of an expected format was not given because it has not been collected	synthetic construct	Information does not exist as the sample represents an ab-initio synthetic construct.
			lab stock	Information was not collected as the sample represents a cultured cell line or model organism under long-term lab control.
			third party data	Information does not exist as the metadata was not collected or reported in records predating the 2023 agreement. For use in Third Party data submissions.
missing	not provided	information of an expected format was not given, a value may be given at the later stage	data agreement established pre-2023	Data agreements were established before the 2023 INSDC standard and metadata can not be provided. A value may be given at a later stage
missing	restricted access	information exists but can not be released openly because of privacy concerns	endangered species	Information can not be reported as the target organism is endangered e.g. on the IUCN red-list
			human-identifiable	Information can not be reported as the metadata would make the sample human-identifiable.

Usage of INSDC Missing Value Reporting Terms

Please use the above standardised missing value vocabulary **only if a true value of an expected format for a mandatory field is missing**. If a true value is missing for a **recommended** or an **optional** field, then these fields should not be used for reporting at all. When reporting a missing mandatory field, the eight granular **‘reporting level’** terms need to be preceded with the term *missing*: to declare both the absence of a true value as well as the reason.

Example of usage:

geographic location (country and/or sea): *missing: data agreement-established pre-2023*

collection date: *missing: control sample*

geographic location (country and/or sea): *missing: human-identifiable*