MacaqueNet Glossary

species

- **species_common** - the common name for each unique macaque species, as specified by the IUCN (International Union for Conservation of Nature).
- **species_scientific** - the scientific name for each unique macaque species. This is the last part of the Latin name (specific epithet), not including the genus name. Referred to as "species" internally.
- **species_id** - a unique number randomly assigned to each unique macaque species.

research site

- **research_site** - the name of the research centre, zoo, animal park, or field site at which the data was collected and by which it is known in the literature.
- **site_id** - a unique number randomly assigned to each unique research site where the data was collected.
- **country_site** - the country in which each research site is located.
- **latitude_site** - the latitudinal coordinates for each unique research site in degrees.
- **longitude_site** - the longitudinal coordinates for each unique research site in degrees.

population

- **pop_id** - a unique number randomly assigned for each unique population, defined as a unique combination of species and research site.

group

- **group_name** - the name of the group within a population on which the data was collected and by which it is known in the literature. Group_name is consistently used for the same social group, even if different names were given by different teams or in different data collection periods. If only one group exists within a population, group_name is 'onegrouponly'. Internally, two versions exist: group_name, which is lowercase and contains no spaces, and group_name_display, which is formatted as it appears in the literature.
- **other_names** - alternative names used in the literature for a given group. If there are no other names for a unique group, the other_names column remains blank.
- **group_id** - a unique number randomly assigned to each unique social group.
- **range_management** - the extent to which a group's range is managed. Can be one of three options: "wild" (animals can move freely between different groups or populations, without range restrictions), "free-ranging" (animals can move freely between different groups or populations, but are confined to a specific area intentionally by humans), or "captive" (animals have restricted ranges and cannot move freely or disperse between different groups or populations).
- **medical_intervention** - whether a group receives medical interventions in the form of reproductive control and/or treatment of injury or disease as part of the population's management plan. Can be one of two options: "1" (routine medical interventions) or "0" (no routine medical interventions).
- **food_provisioned** - whether humans routinely and intentionally provide a group with food. Can be one of two options: "1" (food provisioned) or "0" (not food provisioned).
- **human_contact** - whether a group is in regular contact with non-researcher humans, such as groups whose ranges overlap with human settlements or groups regularly visited by tourists. Can be one of two options: "1" (regular contact with non-researcher humans) or "0" (no regular contact with non-researcher humans).

group_period

- **start_month** - the month in which data collection started for each study period, denoted as the first 3 letters of the month name.
- **start_year** - the year in which data collection started for each study period, denoted in 4-digit format. Referred to as "year" internally.
- **end_month** - the month in which data collection ended for each unique period, denoted as the first 3 letters of the month name.
- **end_year** - the year in which data collection ended for each study period, denoted in 4-digit format.
- **group_period_id** - a unique number randomly assigned to each unique group-period, defined as a unique combination of social group and study period.
- **group_size** - the total number of individuals in each social group during the period of data collection. This value includes all group members: (sub)adults, juveniles, and infants. If group size varies across the period of data collection, the largest group size is noted. If the exact group size is not known, the group_size column remains blank.
- **adult_males** - the total number of adult males in the group at the time data were collected. This value includes males who were or became adults during the study period. If the number of adult males in the group varies across the period of data collection, the largest number of adult males is noted.
- **adult_females** - the total number of adult females in the group at the time data were collected. This value includes females who were or became adults during the study period. If the number of adult females in the group varies across the period of data collection, the largest number of adult females is noted.

unit

- **unit_id** - a unique number randomly assigned to each unique unit, defined as all subject and interaction data on a specific set of subjects from the same group-period. Most group-periods contain only one unit, but multiple units can exist where data for males and females were shared separately or if different subjects were observed using different sampling methods.
- **unit_code** - code for each unit, formatted as 'species_source_group.name_year_within.year.period'.
- **nr_ama_subjects** - the number of adult male subjects for which behaviour was recorded.
- **nr_afe_subjects** - the number of adult female subjects for which behaviour was recorded.
- **all_ama_obs** - whether or not all adult males who were present in the group during the period of data collection were observed. Can be one of three options: "1" (yes, all adult males in the group were observed), "0" (no, not all adult males in the group were observed), or "NA" (there are no adult male subjects for this unit).
- **all_afe_obs** - whether or not all adult females who were present in the group during the period of data collection were observed. Can be one of two options: "1" (yes, all adult females in the group were observed), "0" (no, not all adult females in the group were observed, or "NA" (there are no adult male subjects for this unit).
- **exact_ages_known** - whether the chronological age for most subjects is known. Can be one of two options: "0" (exact ages not known), or "1" (exact ages known for most subjects).
- **filename_matrix** - the name of the file containing each cleaned, standardised socio-metric matrix of behavioural interaction data.
- **filename_subjectdata** - the name of the file containing the subject data for each unit.
- **version** - the version of the MacaqueNet database from which the data were extracted when sharing.

datakey

- **ready_to_deploy** - whether a given sociometric matrix or subject data table is ready to be used. Can be 1 of 2 options: "1" (data has been cleaned, standardised and no crucial issues remain), "0" (data has not been fully cleaned, standardised and crucial issues may remain).
- **crucial_issue_remaining** - whether there are any crucial remaining issues for a given sociometric matrix or subject data table. Can be 1 of 2 options: "0" (no remaining issues), "1" (crucial issues remain).

- **permis_display_web** - whether the data contributors gave permission for their metadata to appear on the database search tool on the MacaqueNet website.
- **macaquenet_id** - a unique number randomly assigned to each socio-metric matrix or subject data table. Referred to as "running_number" internally.
- **group_name_verified** - whether a group_name has been verified by the data contributor.
- **within_year_period** - a chronological marker assigned for units from the same group-period (default "1" for a single dataset). If distinction within the same group-period is needed (e.g., different subjects for different observation protocols), "1a" and "1b" are used to indicate distinct units.
- **folder** - the name of the folder in which the original data (as provided by the data contributor, before going through the standardisation pipeline) is stored within the primary data repository, denoted as the species' scientific name and the last name of the lead researcher who contributed the data to the database.
- **file** - the name of the file containing the original data, as provided by the data contributor to the database.
- **sep** - the format of each original dataframe. Can be one of four options: "tab" (for dataframes shared as tab separated files such as .txt files), "comma" (for dataframe shared as comma separated files, such as .csv files), "excel" (for dataframes shared as Excel sheets) or "manual" (for datafiles that had to be entered manually).
- **sheet** - for data shared as Excel files with multiple sheets. Denotes the name of the sheet containing each dataframe, as provided by the data contributor to the database. If data were not shared as Excel files, the sheet column remains blank.
- **sheet_range** - for primary data that contains multiple dataframes (e.g. behavioural matrices) on a single sheet. Denotes the cell range containing each dataframe. If data were not shared as Excel files, the Sheet column remains blank.
- **category_global** - the broad data category of the dataframe. Can be one of three options: "affi" (affiliative) or "aggr" (aggressive) for behavioural data or "subjectdata" for subject data.
- **data_category** - the behavioural data category. Can be one of six options: "conagg" (aggression that involves contact, e.g. biting, hitting), "nonconagg" (aggression that does not involve contact, e.g. threatening, chasing), "allagg" (if aggression type is not specified or if both contact and non-contact aggression are combined into one matrix), "bodycon" (being in body contact, e.g. sitting in contact, embracing), "groom" (social grooming) or "prox" (time spent in spatial proximity). Referred to as "cat_sub" internally.
- **category_detail** - provides information on the exact behaviour types within each affiliative and aggressive data_category.
- **focal_data** - whether data were collected via focal animal sampling or group sampling. Can be one of two options: "1" (focal animal sampling) or "0" (group sampling).
- **sampling_method** - the sampling method used to record the behaviour. Can be one of two options: "discrete" (recording at regular time intervals whether the behaviour of interest is occurring, sometimes also called 'time sampling', 'instantaneous sampling' or 'scan sampling') or "continuous" (recording all instances of a behaviour continuously, sometimes also called 'all occurrence sampling').
- **data_type** - the format of the data in a socio-metric matrix. Can be one of two options: "count" (the number of occurrences of a behaviour, if the behaviour was recorded continuously, or the number of discrete samples in which the behaviour was occurring, if the behaviour was recorded discretely) or "duration" (the length of time a behaviour occurred for, in seconds).
- **sampling_interval** - for data collected using discrete sampling. Denotes the interval between discrete samples, in seconds. If data were collected continuously, the interval column remains blank. Referred to as "interval" internally. For units with multiple discrete sampling protocols, there are 2 sampling_interval columns: "interval" and "interval_2".
- **obseff_column** - refers to which observation effort column in subject_data is relevant for the given sociometric matrix. Can be one of four options: "obseff_samples"/"obseff_samples_2" (observation effort is expressed as a number of time samples) or "obseff_duration" "obseff_duration_2" (observation effort is expressed as a duration).
- **corrected_obseff** - whether the data in the sociometric matrix are shared already corrected for observation effort.

- **has_obseff** - whether individual observation effort is available for each subject of a socio-metric matrix.
- **sum_dyad_obseff** - whether the individual observation efforts should be summed to obtain the dyadic observation effort. Can be one of 3 options: "1" (dyadic observation effort can be calculated as the sum of both individuals' observation efforts), "0" (dyadic observation effort should not be the sum of the individuals' efforts, but rather the individual observation effort, such as in group scans where all individuals are present), "-1" (dyadic observation effort cannot be determined from individual efforts, such as in group scans where it is unknown which individuals were present for each scan).
- **data_symmetric** - whether a sociometric matrix is symmetrical or not. Can be one of two options: "0" (the matrix contains directed behaviour with a clear actor and receiver, and is not symmetrical) or "1" (the matrix contains undirected behaviour, and is symmetrical).
- **required_transposing** - whether the socio-metric matrix needed to be transposed during the cleaning and standarising process.
- **notes** - extra qualitative information about each sociometric matrix or subject file.

subject_data

- **subject** - the identification name of each subject from which data were collected for a given unit, as specified by the data contributor.
- **sex** - the sex of each subject. Can be one of two options: "male" or "female".
- **age** - the age of each subject in years, rounded to the nearest integer. The age column may be left blank if a subject's exact age is not known, but age category is known.
- **age_category** - whether the subject is an adult, subadult, juvenile, or infant. Age categories are based on the age cutoffs assigned by the data contributor.
- **partner_only** - whether the subject is an individual of primary interest, or whether data was only recorded on the subject when it was a partner to an individual of primary interest. Can be one of two options: "0" (the subject is of primary interest), or "1" (the subject is not of primary interest).
- **obseff_samples** - for data collected using discrete sampling only. Denotes the observation effort for each unique subject, expressed as a number of discrete samples. For units with multiple discrete sampling protocols, there are 2 obseff_samples columns: "obseff_samples" and "obseff_samples_2".
- **obseff_duration** - for data collected using continuous sampling only. Denotes the observation effort for each unique subject, expressed as a duration in hours. For units with multiple discrete continuous sampling protocols, there are 2 obseff_duration columns: "obseff_duration" and "obseff_duration_2".

permissions

- **metadata_access** - the level of access to the metadata of each group-period, as selected by the data contributor. Can be one of two options: "open" (open access to the metadata) or "onrequest" (access to the metadata can be requested by a researcher wishing to utilise the data for their project).
- **data_access** - the level of access to the behaviour or subject data for each group, as selected by the data contributor. Can be one of two options: "open" (open access to the data) or "onrequest" (access to the data can be requested by a researcher wishing to utilise the data for their project).
- **permission_id** - denotes the access levels for each group-period. Can be one of four options, representing each unique combination of metadata_access and data_access: "1" (open access to both the metadata and the data), "2" (both metadata and data are only available on request), "3" (open access to metadata, but data only available on request).

team

- **team_id** - a unique number randomly assigned to each unique team of researchers who contributed data. A single researcher may be a member of multiple teams. The decision as to who is considered a member of the team is determined by the data contributor.
- **lead_contact** - whether or not a researcher is the lead contact for that team. Can be one of two options: "1" (yes) or "0" (no).

researcher

- **researcher_first** - the first name of each researcher who contributed data to the database.
- **researcher_last** - the last name of each researcher who contributed data to the database. Referred to as "source" internally.
- **researcher_id** - a unique number randomly assigned to each unique researcher who contributed data to the database. Each member of a team who contributed data has a unique researcher_id.
- **researcher_email** - the current email address of each researcher who contributed data to the database.

affiliations

- **institute** - the name of the institution(s) that each researcher who contributed data to the database is currently affiliated with.
- **institute_id** - a unique number randomly assigned to each unique institute that each researcher who contributed data to the database is currently affiliated with.
- **country_institute** - the country in which each unique institution to which each data contributor is affiliated is located.
- **latitude_institute** - the latitudinal coordinates for each unique institution to which each data contributor is affiliated, given in degrees and rounded to the nearest integer.
- **longitude_institute** - the longitudinal coordinates for each unique institution to which each data contributor is affiliated, given in degrees and rounded to the nearest integer.

research_project

- **project_name** - the full name of each unique project using the MacaqueNet database, as it is named on the MacaqueNet website.
- **project_id** - a unique number randomly assigned to each unique research project that is using the MacaqueNet database.
- **project_contact_first** - the first name of the person who serves as the primary point of contact for each research project.
- **project_contact_last** - the last name of the corresponding person who serves as the primary point of contact for each research project.
- **project_contact_email** - the email address of the corresponding person for each research project.

project_data

- **access_requested** - whether the socio-metric matrix has been requested to be used in each specific research project. Can be one of 2 options: "1" (permission has been requested) or "0" (permission has not been requested).
- **access_granted** - whether the data contributor has granted permission for the socio-metric matrix to be used in each specific research project. Can be one of 2 options: "1" (permission has been granted) or "0" (permission has not been granted).