# MEGA: Merger Graph Format

This document details the data format for the graph files produced by MEGA. It is a **work in progress** and therefore **subject to change**.

Any dataset of length $N_{host}$, $N_{sub}$, $N_{gen}$ and $N_{graph}$ is ordered such that the index is the corresponding ID (all IDs start at 0). For example, to get data for graph 3 from an $N_{graph}$ length array (`arr`) you would simply index it thusly:

```
arr[3]
```

Any dataset of length $N_{prog}$, $N_{desc}$, $N_{subprog}$, and $N_{subdesc}$ contain the data for progenitors and descendents and are of ambiguous length. Due to this there are associated pointer (start_index) and length (nprog, ndesc) arrays. This is also true when trying to extract individual generations from the $N_{host}$ and $N_{sub}$ arrays using the generation start and length arrays. Data can be extracted from these using:

```
arr[start: start + length]
```

## Root:

| Groups | |
|---|---|
| Key | Description |
| `Header` | Contains Metadata for the simulation. |
| `N` | The groups containing the data for each individual graph. N runs from 0 - $N_{graph}$-1. |

| Datasets | | | |
|---|---|---|---|
| Key | Description | Units | Type/ Shape |
| `graph_lengths` | The length of each graph. | None | int[]/ ($N_{graph}$, ) |
| `root_nparts` | The number of particles in the most massive host halo in the root generation of each graph (lowest redshift generation). | None | int[]/ ($N_{graph}$, ) |
| `nhalos_in_graph` | The number of halos in each graph. | None | int[]/ ($N_{graph}$, ) |
| `sub_graph_lengths` | The length of all subhalo graphs. | None | int[]/ ($N_{graph}$, ) |
| `sub_nhalos_in_graph` | The number of subhalos in each graph. | None | int[]/ ($N_{graph}$, ) |

| Attributes | | | |
|---|---|---|---|
| Key | Description | Units | Type |
| | | | |
| | | | |

## Header:

| Datasets | | | |
|---|---|---|---|
| Key | Description | Units | Type/Shape |
| | | | |
| | | | |

| Attributes | | | |
|---|---|---|---|
| Key | Description | Units | Type |
| `part_mass` | The dark matter particle mass. | M_sun | float |
| | | | |

## Graphs (N):

| Attributes | | | |
|---|---|---|---|
| Key | Description | Units | Type |
| length | The length of this graph, i.e. the number of generations | None | int |
| root_mass | The number of particles in the most massive halo in the root generation (lowest redshift generation) | None | int |
| nhalos_in_graph | The number of halos in this graph | None | int |
| sub_length | The length of this subhalo graph, i.e. the number of generations containing subhalos | None | int |
| sub_root_mass | The number of particles in the most massive subhalo in the root generation (lowest redshift generation) | None | int |
| sub_nhalos_in_graph | The number of subhalos in this graph | None | int |

| Datasets | | | |
|---|---|---|---|
| Key | Description | Units | Type / Shape |
| graph_halo_ids | The internal graph halo ID assigned to a halo. NOTE: These differ from the halo catalog. These run from 0 - $N_{host}$ with the index equal to the value. | None | int[] / ($N_{host}$, ) |
| halo_catalog_halo_ids | The halo catalog ID assigned to each halo. | None | int[] / ($N_{host}$, ) |
| snapshots | The index of the snapshot in the snapshot text file dictated in the param file, for each halo. | None | int[] / ($N_{host}$, ) |
| redshifts | The redshift for each halo in the graph. | None | float[] / ($N_{host}$, ) |
| generation_id | The ID (or number) associated with each generation. Counting starts from the earliest snapshot. | None | int[] / ($N_{gen}$, ) |
| nparts | The number of dark matter particles in each halo. | None | int[] / ($N_{host}$, ) |

| | | | |
|---|---|---|---|
| mean_pos | The mean position of the particles in the halo. | cMpc | float[] / ($N_{host}$, 3) |
| generation_start_index | The starting index (pointer) for each host halo generation. | None | int[] / ($N_{gen}$, ) |
| generation_length | The number of halos in each generation. | None | int[] / ($N_{gen}$, ) |
| nprog | The number of progenitors for each halo. | None | int[] / ($N_{host}$, ) |
| ndesc | The number of descendents for each halo. | None | int[] / ($N_{host}$, ) |
| prog_start_index | The starting index (pointer) for each halo's entries in all progenitor halo arrays (i.e. direct_prog_ids, direct_prog_contribution, etc.). Entries containing 2**30 have no descendants. | None | int[] / ($N_{host}$, ) |
| desc_start_index | The starting index (pointer) for each halo's entries in all descendant halo arrays(i.e. direct_desc_ids, direct_desc_contribution, etc.). Entries containing 2**30 have no descendants. | None | int[] / ($N_{host}$, ) |
| direct_prog_ids | The progenitor halo IDs, extracted using prog_start_index and nprog. | None | int[] / ($N_{prog}$, ) |
| direct_desc_ids | The descendent halo IDs, extracted using desc_start_index and ndesc. | None | int[] / ($N_{desc}$, ) |
| direct_prog_contribution | The number of dark matter particles contributed **by** each direct progenitor **to** the halo. | None | int[] / ($N_{prog}$, ) |
| direct_desc_contribution | The number of dark matter particles contributed **to** each direct descendent **from** the halo. | None | int[] / ($N_{desc}$, ) |
| sub_graph_halo_ids | The internal graph subhalo ID assigned to a subhalo. NOTE: These differ from the halo catalog. These run from 0 - $N_{sut}$ with the index equal to the value. | None | int[] / ($N_{subt}$, ) |
| subhalo_catalog_halo_ids | The subhalo catalog ID assigned to each subhalo. | None | int[] / ($N_{sub}$, ) |
| sub_snapshots | The index of the snapshot in the snapshot text file dictated in the param file, for each subhalo. | None | int[] / ($N_{sub}$, ) |

| | | | |
|---|---|---|---|
| `sub_redshifts` | The redshift for each subhalo in the graph. | None | float[] / ($N_{sub}$, ) |
| `sub_generation_id` | The ID (or number) associated with each generation. Counting starts from the earliest snapshot. | None | int[] / ($N_{gen}$, ) |
| `sub_nparts` | The number of dark matter particles in each halo. | None | int[] / ($N_{subt}$, ) |
| `sub_mean_pos` | The mean position of the particles in the subhalo. | cMpc | float[] / ($N_{sub}$, 3) |
| `sub_generation_start_index` | The starting index (pointer) for each subhalo generation. | None | int[] / ($N_{gen}$, ) |
| `sub_generation_length` | The number of subhalos in each generation. | None | int[] / ($N_{gen}$, ) |
| `sub_nprog` | The number of progenitors for each subhalo. | None | int[] / ($N_{subt}$, ) |
| `sub_ndesc` | The number of descendents for each subhalo. | None | int[] / ($N_{subt}$, ) |
| `sub_prog_start_index` | The starting index (pointer) for each subhalo's entries in all progenitor subhalo arrays (i.e. `sub_direct_prog_ids`, `sub_direct_prog_contribution`, etc.). Entries containing 2**30 have no descendants. | None | int[] / ($N_{subt}$, ) |
| `sub_desc_start_index` | The starting index (pointer) for each subhalo's entries in all descendant subhalo arrays (i.e. `sub_direct_desc_ids`, `sub_direct_desc_contribution`, etc.). Entries containing 2**30 have no descendants. | None | int[] / ($N_{subt}$, ) |
| `sub_direct_prog_ids` | The progenitor subhalo IDs, extracted using `sub_prog_start_index` and `sub_nprog`. | None | int[] / ($N_{subprog}$, ) |
| `sub_direct_desc_ids` | The descendent subhalo IDs, extracted using `sub_desc_start_index` and `sub_ndesc`. | None | int[] / ($N_{subdesc}$, ) |
| `sub_direct_prog_contribution` | The number of dark matter particles contributed **by** each direct progenitor **to** the subhalo. | None | int[] / ($N_{subprog}$, ) |
| `sub_direct_desc_contribution` | The number of dark matter particles contributed **to** each direct descendent **from** the subhalo. | None | int[] / ($N_{subdesc}$, ) |

| | | | |
|---|---|---|---|
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |