

OTel Entity Data Model - Alternative

This is an alternative proposal to [OpenTelemetry Entity Data Model](#)

.

Status: Draft

Why

The initial approach does not specify whether the keys of the composite id are always the same for a given type of entity or if they can vary from one environment to another. If the keys are fixed for a type of entity, we lose flexibility (e.g., the way to identify a service can be different from one cloud provider to another; a service can be regional or global; a service can be multi-tenant or not, etc.). If the keys can change from one environment to another, it makes the implementation of telemetry data consumers more complex and costly. For example, to perform an aggregation by entity, one would have to do something like: 1) sort the key/value pairs by the keys, 2) create a unique ID from this sorted set.

[More TBD]

Observation

Identifying entities makes them navigable, allows the definition of relationships between entities, and enables the aggregation of signals relative to a specific entity. These three properties do not require knowing the exact list of immutable attributes that form the entity's identification.

Proposal

We can achieve the same result with only one field, something like an URI type consisting of two parts separated by `:` (if the URI concept is not suitable, we could use two fields instead of one). The first part is the unique global identifier of the organization or system issuing this entity ID. The second part is the identifier of the entity which is guaranteed by the organization or system to be unique locally and persistent throughout the entity's lifecycle. In the absence of a centralized database maintaining the global uniqueness of the first part of this ID, we could use the domain name of the organization that issued this entity.

For the consumers of these entity IDs, things are simple. With no ID canonization process to perform, they know by convention that there is only one opaque identification field that can be directly used to link entities, to aggregate signals relative to this entity, or for any other reason.

The attributes of the entity remain present to provide a list of key/value pairs characterizing an entity but are not directly used as a unique identifier. These attributes can be used for filtering, sorting, aggregations, etc.

If it is necessary for consumers to know the list of attributes that participate in the identification of an entity for reasons of data governance/data lineage or other, I propose using the concept of an application telemetry schema to allow this information to be exposed at the schema level instead of within the data stream itself.