PHES-ODM-Documentation

Table of contents

Introduction

High-level overview of the ODM

What is the ODM

Contributing

Acknowledgement

Getting to know the ODM

What are "Parts" and "Part Types"?

Within the Public Health and Environmental Surveillance Open Data Model (PHES-ODM, hereafter referred to as simply "the ODM"), every piece or component of the model structure is called a "part". Regardless of function, all of the tables, headers, and categorical inputs are all parts. This creates a foundational term so that the ODM and its structure can be discussed with greater ease by users and designers.

Because the term "parts" refers to every component of the ODM, "part types" are a way to differentiate between parts which have different functions and structures. The three most important part types are:

- Method Part Type
- Measure Part Type
- Attribute Part Type

Please see the parts reference document for more details, but essentially methods allow the recording of methods in environmental surveillance and testing, while measures allow the recording of observations and results, and attributes capture metadata.

The remaining 20 part types (for a total of 23 part types) are:

- Aggregation Part Type
- Aggregation Scale Part Type
- Aggregation Set Part Type
- Category Part Type
- Category Set Part Type
- Class Part Type
- Compartment Part Type
- Compartment Set Part Type
- Dictionary Part Type
- Domain Part Type
- Group Part Type
- Missingness Part Type
- Nomenclature Part Type
- Quality Part Type

- Quality Set Part Type
- Specimen Part Type
- Specimen Set Part Type
- Table Part Type
- Unit Part Type
- Unit Set Part Type

For information and details about all the part types, please see the parts reference document.

"Sets" and the rationale behind sets

"Sets" are a way of grouping together different possible categorical inputs within the ODM. The purpose behind sets is to group options together so that users only need to check a few options from a drop-down menu, rather than scroll through a long list. Sets are also designed so that a single part can be used in multiple sets, which avoids the need to create multiple versions of this part for these different use contexts. For example, if you're taking a measure of concentration, that measure will likely populate the unit choice drop-down list with units from the "standard concentration unit set". Units in this set include Milligrams per litre, parts per million, and Percent Primary Sludge. Similarly, if one were measuring the amount of oxygen in a wastewater sample they would be offered units from the "dissolved gas unit set". The units in this set include parts per million, and Milligrams per litre. You see here how Milligrams per litre is used twice across the sets, but exists still as a single part in the parts list. Unit sets are only one type of set, with the others being:

- Aggregation sets
- Compartment sets
- Quality sets
- Specimen sets

Details about these sets are their similar - but unique - applications can be found in the parts reference document. Importantly, "category sets" are not considered a "set" like these others. See the section below for an explanation of this difference.

Why category sets are treated differently

One of the reasons why "sets" are set up and organized the way that they are, is to allow for the re-use of parts in multiple sets. Category sets are, however, an exception to this rule because the categories that make up a category set are used in only one set. Furthermore, the general sets are needed for almost every measure or method (which require unit, aggregation, quality, and compartment information), while category sets are used only in specific circumstances for specific fields. These are things such as collection metadata (example: Sample collection

category set), dictionary metadata and model architecture (example: Data Type Category Set), or more detailed types of certain methods (example: Nucleic Acid Extraction Category Set). Category sets are also a distinct part type, and more information about them can be found in the parts reference document.

The "what" and "why" of Specimen IDs

Specimen IDs indicate the level at which a measure is being done. This can be: a site measure, ie. a measure at a site such as temperature or weather; a sample measure, ie. a laboratory measurement on a sample, such as gene copies of SARS-CoV-2 per mL; a person measure, ie. a measure of something at the level of an individual person, such as a blood pressure reading; or a population measure, ie. an aggregate measure of a population, such as the number of confirmed cases of a given illness. While there is currently nothing in version 2.0 of the ODM with uses the person specimen ID explicitly, it has been included to allow for the possibility of storing individual-level data in a future version. The reason specimen IDs were created was so that there could be a single manner in which measures and methods are recorded, regardless of the level at which they were performed, while still maintaining the possibility to collect site, sample, and population measures as a distinct types of metadata.

The "what" and "why" of Groups and Classes

Groups and classes, similar to sets, are ways of grouping together different measures within the ODM. Given that the ODM aims to be as robust as possible and has a very long list of possible measures, groups and classes were designed to give shorter lists of measures in the drop down lists by specifying details about the kind of measure that a user is recording. Groups and classes can work together to further specify what kinds of measures are being reported. For example, the group `sarsCov2` contains many measures, but by specifying that the class is an `allele` or `variant`, the list of possible measures is pared down. Alternatively, class can be said to be non-applicable, paring down the measures in the `sarsCov2` group to only be unspecified measures of the quantity of the virus.

Tables and table types within the ODM

Within the ODM there are three types of tables: program description tables, results tables, and look-up tables. These table types exist to differentiate the function of these tables and to highlight these differences to users. The different tables take different types of inputs and maintenance from users, so understanding the differences can be important.

Program description tables

Program description tables (represented in yellow in the ERD) are tables used to record metadata on the organizations, locations, methods, and appurtenance. These tables help to describe surveillance and testing programs, and are intended to be updated infrequently.

Results tables

Results tables (represented in blue in the ERD) are the tables used to record details on samples and measures. These tables record the main outcomes data and are updated daily, if not more frequently.

Look-up tables

Look-up tables (represented in green in the ERD) are the tables that are pre-programmed and pre-populated in the ODM. These hold information on sets, all parts, languages, and translation abilities. These are only updated by the ODM team in version updates.

Measures, Methods, and Attributes: Key parts

While there are many part types, there are three main part types users should be most familiar with: measures, methods, and attributes. These three have parallels with the three table types, and the differences and details of these parts are useful to understand.

Measures

Measures are actually types of measures that can be performed. These can range from temperature to the number of gene copies in a sample. The measure, or `measureID`, can be selected from a drop down in the templates and it specifies the kind of measure you intend to record. The actual value of the measure is then recorded in the `value` field of the `measures` table, with units and aggregation specified in the `unitID` and `aggregationID` fields.

Methods

Similar to measures, methods are types of methods that can be performed to accomplish a measurement. These can be diverse, ranging from incubation, qPCR, or nucleic acid extraction. The method itself (`methodID`) is selected from a drop down in the templates, specifying in general terms the kind of method the user wishes to record. From there, the `value` field of the `methodSteps` table can be populated by one of the inputs from the category set associated with that `methodID`. This provides a higher level of detail for a given method. For example, the `methodID` might be `solidSep` for solid separation of a sample. The value field might then be populated with `cent` for centrifugation, indicating more details about how the sample settling was done.

Attributes

Attributes are the largest category of part type, as these refer to any most of the fields in the ODM. Attributes are fields for metadata within the ODM and range greatly in the use. They comprise everything from collection dates, to sampling period, to names.

Data quality and reportability

Within the ODM there is a `qualityFlag` field present in both the `samples` and `measures` tables. This field serves to highlight whether or not there is a quality issues with the sample or the measure. The quality flag also allows a user to specify to type of quality issue. This is managed through the use of quality sets, which are the sets that contain the possible quality flags for a given measure or for samples. The idea is that this provides data on any issues with a sample and an indication of the nature of that issue. Having data about the nature and presence of a quality issue is often not sufficient for decision makers who are trying to use and interpret the data. As such, there is also a `reportable` field with is a Boolean indicator of whether or not data can or should be reported or included in final reports and decisions.

Time periods for samples and measures

Fo measures, there is a field `aDateStart` and `aDateEnd` which specify the date and time that an analysis was begun and finished. This allows for the recording of greater detail around timelines for especially long, multi-day analyses. For shorter analyses, which will likely make up the bulk of reported measures, the same dat can input in both fields. The idea is to have the reporting tables for measures be as robust as possible to allow for various kinds of timelines.

For samples, there is `collDT`, `collDTStart`, and `collDTEnd`. The first field is for the collection date and time of a single grab sample, so start and end are not necessary and the other two

fields can be left blank. For composite or pooled sample, the start and end date and time for collection is crucial information to know. When these two fields are populated, the `collDT`field can be left blank.

Dates within the ODM

There are a number of date fields within the ODM which all serve different purposes. Collection datetime (`collDT`) is the date a sample was collected, used only for grab samples. This field is left blank if the collection datetime start and end fields (`collDTstart` and `collDTEnd`) have been populated instead. Inversely, collection datetime start and end fields should be left blank if the collection datetime filed has been populated. Similarly, analysis datetime start and analysis datetime end (`aDateStart` and `aDateEnd`) are used to report the date and or timeline of an analysis for a measure.

Date fields that are more related to data processing and labratory infrastructure are: the last edited date (`lastEdited`) which indicates the last time a table, or a measure or sample details, were last edited or updated; the sent date (`sentDate`), or date that a sample was sent to the lab from the field; the received date (`recDate`), or the date the sample was recieved in the lab; and the report date (`repDate`), or the date that the analysis results or measures were reported.These give a greater indication of how up to date data is, but also on the speed and efficiency of the pipeline between sampling and reported results. Dates should be reported in day/month/year format to accommodate the most popular global convention.

Translation and language capabilities

The default language of the ODM is English, but French translations of all descriptive elements of the data model dictionary are also available. As additional nation states and partners adopt the ODM, we anticipate that these fields will be translated into other languages as well. The translation capacities are managed through the language look-up table (`languageLUs`), the translation look-up table (`translationLUs`), and the parts table (`partsLUs`). When a translation for a given part is not available, the dictionary will default to the English term.The language look-up table stores linguistic and classification codes for spoken human languages, with the most recent ISO639 code being the language ID (`languageID`). In the translation table, the language ID is paired with every part ID (`partID`), along with the label (`partLabel`), description (`partDesc`), and instructions (`partInstr`) translated to that language. The part IDs are linked to the full parts list which otherwise contains metadata that is coded in variables and requires no further translation.

Questions and online community

If any users have additional questions or issues with the ODM, we invite them to check out our Discourse Page for discussion boards and community support. For larger issues, and to ask the ODM team to add additional fields or variables into the model, we encourage users to visit the project's GitHub repository and to create and issue there so that a team member can respond.

How to use the ODM

The how-to guides here. If needed we can group into chapters such as: - how to use the report table templates - how to request a new measure - how to validate ODM data (link to that web documentation)

Tutorials

If needed, we can have chapters with differen tutourials

Part I

Reference guide

The book section has the parts and other lists. The chapters are organized so that there can be persistent links for the look-up tables (parts, sets, language, translations.)

Parts

The list of parts from alphabetical order of part ID.

Ctreat the list so that there is a persistent link for each part. i.e.

Sets

The list of sets in aphabetical order of setsID

Languages

The language table by langID.

Translations

The list of translations by languageID.

Tables

The list of tables in the three main groups (partType = tables, class = results, prgramDesc, lookup).

Report tables

Program Description tables

Look-up tables

partTypes

A list of partTypes, grouped by partType. Group by partType, group and class. ie. for list measures by group and classes within groups.

Aggregations

Attributes

Categories

Category sets

Changes column

Classes

Compartments

Data types

Domains

Groups

Measures

group #1 of measures

class #1 of measues within group #1

Methods

Missingness

Nomeclature

Part types

Quality measures

Specimens

Tables

Units