|  |  |
| --- | --- |
| PMXT03 General Dataset Format | |
| Scope: | Pharmacometric consulting activities |
| Description: | Template general dataset format |
| Owner: | Managing Director |
| Target audience: | All associates |
| Version: | 1.0 |
| Release date: | 01-Dec-2015 |
| Upon use of this template, this table should be removed. | |

IQM GENERAL DATASET FORMAT

In the context of model-based drug development often many different analyses, or pharmacometric activities, need to be performed within the same project. This includes, but is not limited to, graphical exploration, dose-concentration and concentration–response modeling, where many different biomarkers or endpoints might need to be considered. Requesting a separate modeling dataset for each modeling activity considerably strains the resources of the modeler who prepares these datasets or of the supporting programming group, especially in the case when a certain level of validation is required.

The approach used at IntiQuan is to use a general dataset format for pharmacometric analyses. The IntiQuan Modeling (IQM) Tools have been enabled to work with this format; **however, can also work with any traditional NONMEM or MONOLIX data format – directly, or by conversion of traditional data formats into the general dataset format.**

Requirements for development of this dataset format have been a well-defined structure that is project independent (same across projects, compounds, and indications), independent of the modeling activity to be performed, and independent of the modeling tool to be used. Additionally, the format had to have a minimum level of redundancy and include information that is typically not found within traditional pharmacometric analysis datasets: names for variables, units, etc. This renders the dataset immediately understandable without additional documentation, reduces mistakes, and facilitates project hand-over processes.

The IntiQuan General Data Format is documented in this document. This document can serve as a template for dataset specification. Upon its use, this first page should be removed and the contents below changed accordingly to specify the required data in collaboration with the customer.

# IntiQuan General Dataset Format

* *Columns that are currently un-used by IQM Tools are marked in grey-italic font.*
* Optional columns for IQM Tools are marked in orange.
* All numeric values will require at least 4 significant digits, e.g., 12.12333 will become 12.12, 0.00001234566 will become 0.00001235.
* Since datasets for pharmacometric analyses typically are comma-separated-files, no field in the general dataset is allowed to contain commata.

| **Dataset section** | **Column Name** | **Column Content** | **Description** | **Type** | **Modeler’s input** |
| --- | --- | --- | --- | --- | --- |
| **General** | **IXGDF** | Index of record in dataset. Starting from 1, then 2,3,… until last record/row number | This index is kept during post processing of the general dataset format. When rows/records are removed the kept ones will keep the index in IXGDF. In this way contents of derived datasets can always be compared to contents of the general dataset format. | Numeric | No |
| **IGNORE** | Reason/comment related to exclusion of the observation or dose from the analysis | Populated from a set of and/or data driven rules.  Example: positive pre-first dose PK observations.  Field can be left empty. | String | Yes |
| **Identification of subject** | **USUBJID** | Unique subject identifier | Populated from the source data.  Field cannot be left empty. | String | No |
| **CENTER** | Center number | Populated from the source data.  Field cannot be left empty. If center number not available, set to 0. | Numeric | No |
| **SUBJECT** | Subject number | Populated from the source data.  Field cannot be left empty. | String | No |
| **IND** | Numeric indication flag | Numeric indication flag (unique for each entry in INDNAME)  This can be used as a categorical covariate. If column not present, IQM Tools will generate this column during the creation of a task specific dataset. Mappings between INDNAME and IND will then be chosen based on alphabetic ordering of INDNAME, starting from 1 and then 2, 3, etc.  If the user would like to impose a mapping between INDNAME and IND, this optional column is useful. | Numeric | Yes |
| **INDNAME** | Indication name | Populated from the protocol or the provided specifications.  Field cannot be left empty. | String | Yes |
| **Study information** | **COMPOUND** | Name of the investigational compound | Populated from the source data or protocol or provided specifications.  Field cannot be left empty. If no compound involved, set to “NONE”. | String | Yes |
| **STUDY** | Short study name/number | Populated from the source data.  Field cannot be left empty. | String | No |
| **STUDYN** | Numeric study flag | Numeric study flag (unique for each entry in STUDY)  This can be used as a categorical covariate. If column not present, IQM Tools will generate this column during the creation of a task specific dataset. Mappings between STUDY and STUDYN will then be chosen based on alphabetic ordering of STUDY, starting from 1 and then 2, 3, etc.  If the user would like to impose a mapping between STUDY and STUDYN, this optional column is useful. | Numeric | Yes |
| **STUDYDES** | Study title, short description | Populated from the protocol title.  Field cannot be left empty. | String | No |
| **PART** | Part of study as defined per protocol | Populated from the source data and/or protocol and only if more than one part is present. For example:  1=first part  2=second part  etc.  Field cannot be left empty. If no parts defined, set to 1. | String | No |
| **EXTENS** | Extension of the core study | Populated from the source data and/or protocol and only if extension is present. For example:  0=core study  1=first extension  2=second extension  etc.  Field cannot be left empty. | Numeric | No |
| **Treatment group information** | **TRTNAME** | Name of actual treatment given to subject | Populated from the source data, protocol or the provided specifications, and can be recoded for harmonization across trials, if needed.  Field cannot be left empty. | String | Yes |
| **TRT** | Numeric treatment flag | Numeric treatment flag (unique for each entry in TRTNAME)  This can be used as a categorical covariate. If column not present, IQM Tools will generate this column during the creation of a task specific dataset. Mappings between TRTNAME and TRT will then be chosen based on alphabetic ordering of TRTNAME, starting from 1 and then 2, 3, etc.  If the user would like to impose a mapping between TRTNAME and TRT, this optional column is useful. | Numeric | Yes |
| **TRTNAMER** | Name of treatment to which subject was randomized | Populated from the source data, protocol or the provided specifications, and can be recoded for harmonization across trials, if needed.  Field cannot be left empty. | String | Yes |
|  | **TRTR** | Numeric randomized treatment flag | Numeric randomized treatment flag (unique for each entry in TRTNAMER)  This can be used as a categorical covariate. If column not present, IQM Tools will generate this column during the creation of a task specific dataset. Mappings between TRTNAMER and TRTR will then be chosen based on alphabetic ordering of TRTNAMER, starting from 1 and then 2, 3, etc.  If the user would like to impose a mapping between TRTNAMER and TRTR, this optional column is useful. | Numeric | Yes |
| **Visit information** | **VISIT** | Visit number | Populated from the source data.  When visit information not available in source data, it could be imputed from:  - Other assessments happening at the same date,  - PK log (in the case of PK with a sample number).  Field can only be left empty in special cases when nominal visits do not make sense. Examples:   * adverse events * comedication | Numeric | No |
| **VISNAME** | Visit name | Populated from the source data.  When missing, it could be derived from the visit number. It is usually missing for adverse events and comedication.  If it is missing, then set to “UNKNOWN”.  Field cannot be left empty. | String | No |
| **BASE** | Flag indicating assessments at baseline | Derived based on VISIT and VISNAME:  =0 for non-baseline visit,  =1 for first baseline visit,  =2 for second baseline visit,  etc.  Field cannot be left empty. | Numeric | No |
| **SCREEN** | Flag indicating assessments at screening | Derived based on VISIT and VISNAME:  =0 for non-screening visit,  =1 for first screening visit,  =2 for second screening visit,  etc.  Field cannot be left empty. | Numeric | No |
| **Event time information** | **TIMEUNIT** | Unit of all Numeric time definitions in the dataset | Populated from the provided specification.  E.g., values could be “hours”, “days”, “weeks”.  Field cannot be left empty. | String | Yes |
| **DATEDAY** | Start date of event | Populated from source data and/or may need to be imputed when it’s partial or missing.  Formatted as DD-MMM-YYYY (Example: 01-JUL-2015)  Field cannot be left empty. If information unknown and cannot be imputed, set to “UNKNOWN”. | String | No |
| **DATETIME** | Start time of event | Populated from source data and/or may need to be imputed when it’s partial or missing, in particular in case of profiles. Formatted as HH:MM (Example: 09:34).  Field cannot be left empty. If information unknown and cannot be imputed, set to “UNKNOWN”. | String | No |
| **DURATION** | Duration of event | Duration of event in same time units as TIMEUNIT.  Derived from start date/time and end date/time of an event.  Populated from source data and/or may need to be imputed when information is partial or missing.  If end time is same as start time (e.g. bolus administration, 0 is a perfectly acceptable value for DURATION).  If event continues post End of Study, set to -1.  Field cannot be left empty. | Numeric | No |
| **NT** | Nominal event time | Planned time of event. Based on protocol, in the time unit defined in TIMEUNIT column. For repeated visits (e.g., 1.001, 2.001), use defined protocol time for the originally scheduled visit. For “End of study” visit, use defined nominal time for end of study – even if subject dropped out and completed End of study visit earlier.  Field cannot be left empty for per protocol planned time dependent assessments.  Field can be left empty, e.g., for adverse events. | Numeric | No |
| **TIME** | Actual time of event relative to first dose administration | In the time unit defined in TIMEUNIT column.  Derived as the difference between DATEDAY/DATETIME of the event and the DATEDAY/DATETIME of the first administration of selected primary study drug.  Field cannot be left empty. | Numeric | No |
| **Event value information** | **TYPENAME** | Unique type of event | Populated from the specification.  For example: “Efficacy Readouts”, “Lab Values”, “Adverse Events”, etc.  Field cannot be left empty. | String | Yes |
| **NAME** | Unique short name of event | Populated from the specification.  For example: “Plasma concentration Compound X”, “Dose Compound X”, etc.  Field cannot be left empty. | String | Yes |
| **VALUE** | Value of event defined by NAME | Applicable to Numeric readouts.  Populated from the source data in the units defined in the UNIT column. Examples for the values: the given dose, the observed PK concentration, or the value of other readouts.  Special cases:  - For adverse events it contains the Grade of the adverse event (and the UNIT column will contain “AE Grade”  - For concomitant medications: the dose in pre-specified unit and frequency (e.g., mg/day)  - For BLOQ records: value 0.  Can only be left empty if VALUETXT is not missing. | Numeric | No |
| **VALUETXT** | Text version of value | Applicable to categorical readouts.  Populated from the source data.  Examples: “Male”, “Female”, “Asian”, etc.  Can only be left empty if VALUE is not missing. | String | No |
| **UNIT** | Unit of the value reported in the VALUE column | Populated/converted from the source data into a reference unit. For same event the same unit has to be used across the dataset. For Labs, Vitals, etc., SI unit is the default. For dimensionless readouts use “NONE”.  Field cannot be left empty. | String | No |
| **ULOQ** | Upper limit of quantification for event defined by NAME | Populated from source data /protocol for PK/LAB assessment only if it is applicable.  Field can be left empty if ULOQ unknown. | Numeric | No |
| **LLOQ** | Lower limit of quantification for event defined by NAME | Populated from source data /protocol for PK/LAB assessment only if it is applicable.  Field can be left empty if LLOQ unknown. | Numeric | No |
| **Dose event additional information** | **ROUTE** | Route of administration | Populated from the source data and/or protocol.  Field is to be left empty for observation events.  Field cannot be left empty for dosing events.  For dosing events the value should be one of the following:  iv, subcut, intramuscular, intraarticular, oral, inhaled, topical, rectal | String | No |
| **INTERVAL** | Interval of dosing | Interval of dosing, if single row should define multiple dosings.  Allows for coding repeated dosing more efficiently.  Populated from source data, protocol or provided specification, in time units defined in the TIMEUNIT column.  Field can be left empty if no multiple dosing events to be specified with a single dosing record. | Numeric | No |
| **NRDOSES** | Number of doses given with the specified interval | Number of doses given with the specified interval, if single row should define multiple dosings,  Allows for coding repeated dosing more efficiently and linked to INTERVAL.  Populated from source data, protocol or provided specification.  Field can be left empty if no multiple dosing events to be specified with a single dosing record. | Numeric | No |
| **Adverse event additional information** | ***SEVERITY*** | *Severity of adverse event* | *Populated from the source data.*  *Field cannot be left empty for adverse events.*  *Field is to be left empty for non-adverse events.* | *Numeric* | *No* |
| ***SERIOUS*** | *Seriousness of adverse event* | *Populated from the source data.*  *Field cannot be left empty for adverse events.*  *Field is to be left empty for non-adverse events.* | *Numeric* | *No* |
| ***DRUGREL*** | *Drug related adverse event or not* | *Populated from the source data*  *Field cannot be left empty for adverse events.*  *Field is to be left empty for non-adverse events.* | *Numeric* | *No* |
| **Additional information** | **COMMENT** | Additional information for the observation/event | Populated from the source data.  Field can be left empty if no comments needed/available for this event. | String | Yes |

# Table of Events/Observations

Note that the table below shows only some examples. The main goal is to use a general dataset format. The actual definitions of TYPE\_NAME and NAME, etc. can be selected more arbitrarily and adapted to the purpose / needs.

| **Type of event** | **TYPE\_NAME** | **Potential label in source data**  **(for selected examples of events)** | **NAME column in dataset**  **(for selected examples of events)** |
| --- | --- | --- | --- |
| **Patient characteristics** | Demographics | BIRTDATE | Birth date |
| GENDER | Gender |
| RACE | Race |
| ETHN | Ethnicity |
| COUNTRY | Country |
| Medical history | DIAGDATE | Date when the diseases was diagnosed |
| TDIAG | Number of years with the disease |
| EXACER | Exacerbation within the last year |
| DIABETES | Diabetes in family |
| PRETREAT | Prior treatments |
| **Population characteristics** | Population | ITT | Intention to treat |
| SAFETY | Safety |
| EFF | Efficacy |
| PP | Per protocol |
| **Dose** | Dose | Dose-Name of the compound | ABC123 dose |
| XYZ123 dose |
| **PK**  **concentration** | PK | ABC123-Serum | Analyte-where it’s measured |
| DEF123-Plasma |
| **Target** | Target | XY\_TOT | Total XY |
| **Lab measurements (chemical entities)** | Lab | GLUCSER | Measurement name-where it’s measured |
| GLUCURI |
| **Vital signs** | Vital signs | SITTDBP | Sitting diastolic blood pressure |
| STANDDBP | Standing diastolic blood pressure |
| TEMPERAT | Body temperature |
| PEF | PEF |
| WEIGHT | Body weight (sometimes assessed continuously and sometimes only at baseline) |
| HEIGHT | Height |
| **Subjective assessments** | Questionnaire | ACQ\_TOT | Name of the questionnaire-Total for total score |
| ACQ\_Q1 | Name of the questionnaire -Question number for specific questions |
| PHADA | Physician’s assessments of disease activity |
| PTADA | Patient’s assessments of disease activity |
| **Response efficacy endpoint**  **Various endpoints that do not fall in any of previous categories** | Response | ACR20 | ACR20 |
| 6MINW | 6 minute walk distance |
| GAITSP | Gait speed |
| XRAY | X-ray |
| **Adverse events** | Adverse event | HEADACHE | Headache |
| CHOLEST | Elevated cholesterol |
| BLOODPRES | High blood pressure |
| DEATH | Death |
| HOSP | Hospitalization |
| **Concomitant use** | Concomitant medications | MTX | Methotrexate |
| STEROIDS | Corticosteroids |
| **Induced events by protocol** | Induced event | FOOD | Food |
| EXERCISE | Exercise |
| ALLERG | Allergen provocation |