

The documentation and data for the animal studies are organized and named in folders according to purpose: Form Illustrations, Data Dictionary, Rules, Data Sets. Form Illustrations are pdf files. All other files are presented in csv format. Each data transfer includes observations that have cleared all quality control parameters (i.e., rules in the Rules file). Data transfers are sequentially numbered, as well as identified by Phase and age for groups of animals, either six or eighteen months (6M or 18M).

Form Illustrations – pdf files of data collection Case Report Forms (CRFs)

- Representations of CRFs illustrate the types, order, and coding of data collected; the data collection form is concatenated with the data entry screen; the data entry screen displays variable names in context
- Current CRF versions are included; differences from former versions involve additions of variables only
- Specific to Animal Specimen Processing: as defined in the PASS protocol, the order of sample processing was different at the three PASS locations; while the CRFs look slightly different, the field names are consistent between CRFs used at these locations; the Iowa CRF has been provided as an example

Data Dictionary – A table describing fields associated with Data Sets

- Fields, in the associated data set, are presented in logical order, by Form version number (versionNbr), sequentially assigned, and typically ordered according to the order of items on the associated case report form (CRF)
- Codes for External Data Release are assigned to each field; these status codes are DMAQC recommendations according to sensitivity: 1=“All” (no confidentiality risks), 2=“Low” (confidentiality risk: minimal or use caution), 3= “High” (confidentiality risk: high, Internal Use ONLY, 4=“System Codes: Internal Use” (system generated linking codes)
- Fields with categorical values have codes defined in the Categorical Values column when applicable
- Calculated or created variables: are identified with a 1 (=yes) in the Calculated/Created column; and include the notation, [C], preceding the text in the Alias (Field Name description) column
- GUID coding: system generated unique alphanumeric 36 character strings, similarly formatted (# characters separated by dashes: 8-4-4-4-12) that provide specific internal links to forms, fields, and form versions
 - formGUID – form or CRF identifier; links to formName
 - fieldGUID – field or variable identifier; links to Field Name
 - versionGUID – form version identifier; links to versionNbr
 - siteGUID – collection site identifier; links to SiteID
- Specific to Specimen Processing – Sample identifiers:
 - labelid – unique numeric identifier associated with the label affixed to the specimen vial
 - labelguid – unique alphanumeric identifier (GUID format) associated with the label affixed to the specimen vial; links to labelid
 - bid – biospecimen identifier, a unique 5 digit code linking all samples collected in the same collection period
 - sampletypeguid – unique alphanumeric identifier (GUID format) associated with sample type; links to sampletypedescription
 - samplenum – sequential number identifying multiple samples from similar specimen source

- aliquotguid – unique alphanumeric identifier (GUID format) associated with sample aliquot; links to aliquotdescription
- Specific to BIC Label Data (Specimen Vial Labeling) – Sample identifiers:
 - labelID – direct link to Specimen Processing labelid
 - vialLabel – a unique 11 digit code, linking vials created from one specimen identified by the labelID; is a composite of BID, Timepoint, Sample Type, and Aliquot number and is the 1-D label on vials shipped from the Biorepository. This field matches the information provided in the shipping manifest from the Biorepository to the CAS.
 - barcode – a unique code, specific to 2-D barcodes used by the Biorepository for specific sample types. This field matches the information provided in the shipping manifest from the Biorepository to the CAS.
- Codes that link between Data Dictionary, Rules, Data Sets: versionNbr, formGUID, fieldGUID, versionGUID
- Codes that link between Data Dictionary, Data Sets: participantGUID, siteGUID, formGUID, versionGUID, versionNbr

Rules – A table describing quality control parameters applied to entered data

- Rules, for the associated data set, are presented; including field specific checks, cross field checks, and cross form checks
- Rules, applied at the time of data entry, are linked to specific form version numbers; post-data entry rules are applied to all data regardless of form version
- Rules are presented in logical order, by form version number (versionNbr), and sequentially assigned; rules applied after data entry are applied simultaneously and are assigned a sequence of 999
- For rules applied after data entry, the fields involved are further defined by adding table names as a suffix to field names; also, forms with multiple records per participant have an additional numeric suffix to indicate the record number queried
- Codes that link between Data Dictionary, Rules, Data Sets: versionNbr, formGUID, fieldGUID, versionGUID

Data Sets – data files with quality control approved observations associated with primarily CRF sourced and keyed data

- data are presented with field names in first row
- fields are ordered according to the associated dictionary
- BIC Label Data: contains data as of the data capture date
 - Note that this dataset contains only those vials present on manifests for vials shipped from the Biorepository to the CAS (and may therefore not include all reported samples processed at the PASS)
- Codes that link between Data Dictionary, Rules, Data Sets: versionNbr, formGUID, fieldGUID, versionGUID
- Codes that link between Data Dictionary, Data Sets: participantGUID, siteGUID, formGUID, versionGUID, versionNbr