

Guide to the leDEA Harmonist Data Toolkit Error Detail Spreadsheet

Each row of the error detail spreadsheet describes one **error record**: one possible inconsistency (error) that was found in the dataset. The goal of the error detail spreadsheet is to provide enough information to locate the observation(s) in the dataset related to each error and also explain the error. One error might involve multiple variables but is attributed to one primary variable, named in the **error_field** column. The value of that variable—the probable source of the error—is listed in the **error** column. The **table** column provides the name of the leDEA table where **error_field** is found. If the error involves more than one variable, those are listed in **error_field2, error2**, etc.

Each **table** in the leDEA DES requires a different set of key variables to uniquely identify one record. To aid data managers in locating specific dataset record(s) related to an error, the identifying variables and their corresponding values are listed in columns **id1_field, id2_field, id3_field** and **id1, id2**, and **id3**.

For errors that involve multiple tables, (for example, comparing BIRTH_D in tblBAS with VIS_D in tblVIS) the secondary table is named in the **crostable** column. The **severity** column ranks the error as a warning (not as serious) or an error (should be corrected, if possible). Text explaining the error is found in **description** (a shorter version of which is found in **category**). General categories of errors are indicated in the **errorCode** column.

One example row of error detail spreadsheet:

PROGRAM	table	id1_field	id1	id2_field	id2	category	error_field	error	description	severity	errorCode	id3_field	id3	error_fi
Program_3	tblVIS	PATIENT	146	VIS_D	01/05/2001	Numeric Format Error	WEIGH	X	Numeric Value Expected for WEIGH	Error	Numeric Format Error			

Error detail spreadsheet column descriptions:

- **PROGRAM**: the PROGRAM associated with the PATIENT in the error record
- **table**: name of table in which the observed **error_field** is found
- **id1_field**: the name of the field that identifies a unique record or observation in **table**
- **id1**: the value of the variable **id1_field** in the error record
- **id2_field/id2** and **id3_field/id3**: If multiple columns are needed to uniquely identify a record in **table**, those are specified in **id2_field/id2** and **id3_field/id3**
- **error_field**: the name of the primary variable related to the possible error
- **error**: the value of **error_field** found to be inconsistent or erroneous
- **category**: a short summary of the error
- **description**: a longer explanation of the error
- **severity**: Warn or Error, depending on how problematic the error is
- If the error involves more variables than just the one named in **error_field**, these are named in columns **error_field2, error_field3**, and **error_field4** and the corresponding values are listed in **error2, error3**, and **error4**.
- **crostable**: if one of the **error_fields** is from a table other than the one named in **table** it is listed here
- **errorCode**: general error category