Common Types of Errors and Warnings for File Uploads

JA Job Aid

Introduction to Data Validation

When CDC evaluates an HIV Prevention program, it is critical to have data that are consistent, complete, and accurate.

Error and warning messages are used in EvaluationWeb as part of validating incoming data from any source. Whether you upload or key enter data, these messages help you manage your records so they meet the basic requirements that both the computer system and the research process need in order to be successful.

When you upload a file, you receive validation messages all at once. When you enter data directly, you receive them one at a time as you enter each record.

Even though there are more than 900 validation rules that cover the data sets CDC collects through EvaluationWeb, the rules fall within a small number of basic themes. This job aid provides you with information on these core concepts that govern how EvaluationWeb reviews your incoming data. Understanding the logic behind the most common validation rules will help you troubleshoot your next upload submission.

Additional Things to remember

- Be mindful of which data set you are working with, what your key variables are, and how many records contain an error or warning.

Other Resources

Basic information on validation rules, including definitions, descriptions, and locations, are well documented in the following guides, videos, and websites:

Documents

- *Uploading Files into EvaluationWeb® User Guide*
  http://evaluationweb.com/helpTest/documents/Uploading%20Files%20to%20Evaluation%20nWeb%20User%20Guide.pdf

  This document provides general knowledge on many topics, from working through the upload process to working with files that have errors.

- *Creating Files to Upload into EvaluationWeb®: A Primer for Programmers*
  http://evaluationweb.com/help/documents/Creating%20Files%20to%20Upload%20Into%20EvaluationWeb%20-%20A%20Primer%20for%20Programmers.pdf

  This document is designed specifically for programmers who need to manage export routines from other data systems that provide a formatted XML or Excel file.

- *NHM&E Data Variables and Values* (commonly referred to as the Data Variable Set or DVS)
  http://evaluationweb.com/help/documents/rptNHME%20HDCBO%20comp%20DVS_3.17.14_Corrected.pdf

Video

- *EvaluationWeb® File Upload Process*

  Available in the EvaluationWeb Video section of the EvaluationWeb Help page, at http://evaluationweb.com/help/.
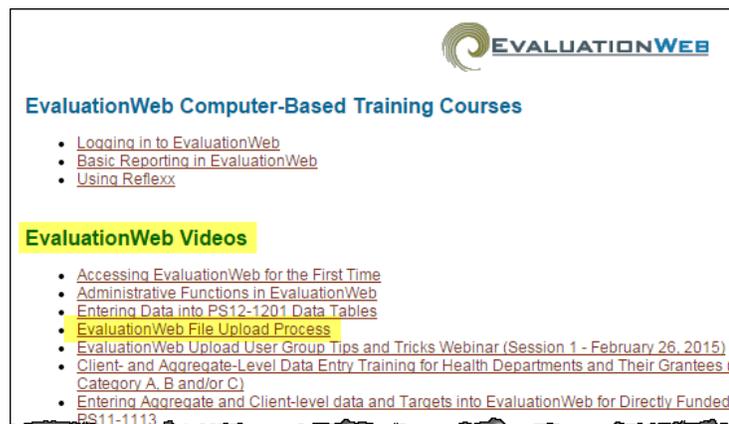
Website

- XML/Excel Upload Resources and Information
  http://schemas.lutherconsulting.com/

  This page contains many reference documents, schema definitions, and sample files.

Questions?

This document is an introduction to some common types of file upload errors and warnings. If you have any questions, we would like to hear from you!

Luther Consulting help desk contact information:

- 1-866-517-6570, option 1
- help@lutherconsulting.com

Five Major Themes

(1) Valid Codes

The general condition that will cause a "...must be a valid code" validation rule error in the File Process Report (FPR) is a variable containing a value not recognized in CDC's NHM&E Data Variable Set (DVS).

An example from an XML file that would cause this error is:

```
...
<birthYear>1934</birthYear>
<ethnicity>E1</ethnicity>
<races>
    <race raceId="1">
        <raceValueCode>2</raceValueCode>
    </race>
        <race raceId="2">
        <raceValueCode>R3</raceValueCode>
        </race>
</races>
...
```

The first **raceValueCode** in the above record is 2, but a quick reference to the DVS shows this value code is not a valid response for Race. Even though the second **raceValueCode** above has a valid value of 'R3', an error message will still be generated because every value must be a valid value for the variable in question. (Note that a value of '2' may be valid for other variables, such as **birthGenderValueCode**, but it is not valid for **raceValueCode**.)

Here's an example specific to the Partner Services Version 2 data set:

```
…
<elicitPartner>
        <timePeriodForRecallInMonths>24</timePeriodForRecallInMonths>
        <totalNumberOfPartnersClaimed>3</totalNumberOfPartnersClaimed>
<totalNumberOfNamedPartners>three</totalNumberOfNamedPartners>
…
```

In this case, you would receive a validation warning message indicating that the **totalNumberOfNamedPartners** needs to be a valid number. However, the value in red above is a number written out English.

The screen shot below shows the DVS entry for the variable G116 (Race) and indicates where to find the information discussed above.

Num	Variable Name		
G116	Race		XSD (Schema) Name: raceValueCode

Value Option: Choose all that apply   **Format Type:** Alpha-Numeric   **Min Length:** 2   ength: 2

Definition: A client's self-reported classification or classifications of the biological heritage with whic   ost closely identify. Standard OMB race codes are applied.

Instructions: Indicate the client's self-reported race(s) using standard OMB race codes. Record all ra   ries that the client reports.

Business rule   HIV Testing: Required, see detailed business rule regarding multiple respo
Partner Services: Required, see detailed business rule regarding multiple
HD Risk Reduction Activities: Required, see detailed business rule regardi
CBO Risk Reduction Activities: Required, see detailed business rule regar
HD Aggregate: Not applicable
CBO Aggregate: Not applicable

Find the **XSD (Schema) Name** from the DVS that matches the element in your file.

Detailed business rule:
Multiple value co   7 or 99. If val   ingle value
may be selected.

Look at the **Code** list. If the code used in your file does not match the code *for that variable* in the DVS, it is invalid.

Code	Value Descript	Value Definition
66		The client was not asked to report his or her race.
77		The client declines or is unwilling to report his or her race.
99		The client reports that he or she is unaware of their race.
R1	nerican Indian	A person having origins in any of the original peoples of North or South America (including Central America), and who maintains tribal affiliation or community attachment.
R2	Asian	A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian Subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.
R3	Black or African American	A person having origins in any of the black racial groups of Africa.
R4	Native Hawaiian or Pacific Islander	A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.
R5	White	A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

Additional things to remember
- The node for a variable (such as Race) must be present within the record whose identifier (form ID or client ID depending on the data set) is listed in the FPR.
- Leaving the node blank or omitting the node completely will not cause an 'invalid code' error message (but it may cause other types of errors or warnings).

(2) Date Format Rules

The general condition that will cause a "... must be a valid date (MM/DD/YYYY)" validation rule error in the FPR is a date variable in a record with any format other than MM/DD/YYYY.

An example that would cause this validation error message to appear in the FPR is:

    ...
    <collectedDateForClient>08/2014</collectedDateForClient>
    ...

In this case, **collectedDateForClient** is incorrectly entered as MM/YYYY, as seen above.

The screen shot below shows the DVS entry for the variable G101 (Date Client Demographic Data Collected) and indicates where to find the information discussed above.

Additional things to remember

- The XML node that describes the variable (for example, <collectedDateForClient>) must be present in the specific record identified in the FPR by Form ID or Client ID.
- Leaving a date variable node blank or omitting the node completely will not cause a date format error (but it may cause other of errors or warnings).
- If there is no date that can be reported for a required variable, EvaluationWeb typically accepts 01/01/1800 as a valid value.

(3) Character Length Rules

The conditions for this type of validation rule message are specific to the variable in question. However, the error or warning message in the FPR will generally give a precise number of characters that a particular value may have.

An example that would cause this validation rule to appear in the FPR is:

```
    ...
    <session>
        <interventionId>This ID is way too long to fit within the 32-character length that the
validation rule and that the NHM&E Data Variables and Values document
specify</interventionId>
        <sessionDate>08/01/2014</sessionDate>
        <siteId>Template_S1</siteId>
    </session>
    ...
```

The above **interventionId** has 148 characters, significantly more than the maximum 32 characters specified in the DVS. Therefore, a validation error message would appear in the FPR for a file containing a record with this **interventionId**.

The screen shot below shows the DVS entry for the variable H01 (Intervention ID) and indicates where to find the information discussed above.

Num	Variable Name			
H01	**Intervention ID**		**XSD (Schema) Name:** InterventionId	
Value Option: N/A		**Format Type:** Alpha-Numeric	**Min Length:** 1	**Max Length:** 32

Definition:	An alpha-numeric identification code used to uniquely identify an intervention.
Instructions:	Enter the alpha-numeric identification code used by your agency to identify the intervention. ...ID is unique for each agency.
Business rule	HIV Testing: Mandatory Partner Services: Required HD Risk Reduction Activities: Mandatory CBO Risk Reduction Activities: Required HD Aggregate: Not applicable CBO Aggregate: Not applicable

Look at the **Min Length** and **Max Length** for this variable in the DVS.

WARNING: Any fifteen character IDs (or longer) entered into an Excel template should be formatted as 'Text.' IDs of this length which are formatted as 'General' or 'Number' will not convert correctly when the file is uploaded to EvaluationWeb.

(4) Variable Dependency Rules and Detailed Business Rules

The general condition that will cause an "If ___ is indicated, then ___" error is when a dependency between multiple variables has not been respected. In other words, a particular response to one variable is required for another variable response to make sense.

A Simple Example

An example that would cause this error to appear in the FPR is:

> …
>
> \<progAnnouncementProgStrategy>89\</progAnnouncementProgStrategy>
>
> \<spfyProgAnnouncementProgStrategy> \</spfyProgAnnouncementProgStrategy>
>
> …

In this example, 89 is given as the response for **progAnnouncementProgStrategy**. Since 89 is an 'Other' option according to the DVS, EvaluationWeb also expects a response for **spfyProgAnnouncementProgStrategy**. However, since no response is given for this variable in the example above, the FPR would generate the following validation rule:

> "If Program Announcement Program Strategy is '89' (Other specify), then Specify Program Announcement Program Strategy is required."

The screen shot below shows the DVS entry for the variable X137-1 (Specify Program Announcement/Strategy) and indicates where to find the information discussed above.

Additional Things to remember

- When you provide a value for **spfyProgAnnouncementProgStrategy**, it must be an exact character match with the name of a program announcement existing in the database. Administrators can check and update the list of program announcements in EvaluationWeb by going to System Utilities > Manage Lists > Manage Program Awards.

A Complex Example

A more complex example that would cause this type of validation rule message to appear in the FPR is:

```
<clientRiskProfiles>
        <clientRiskProfile>
                <dateCollectedForRiskProfile>08/01/2014</dateCollectedForRiskProfile>
                <noClientRiskFactors>5</noClientRiskFactors>
                <vaginalOrAnalSexInLast12Mnths>
                        <withMale>1</withMale>
                        <withFemale>1</withFemale>
                        <withTransgender>1</withTransgender>
                        <vaginalOrAnalSexWithoutCondomPS>1</vaginalOrAnalSexWithoutCondomPS>
                        <vaginalOrAnalSexWithIDUPS>1</vaginalOrAnalSexWithIDUPS>
                </vaginalOrAnalSexInLast12Mnths>
                <injectionDrugUse>1</injectionDrugUse>
                <shareDrugInjectionEquipment>1</shareDrugInjectionEquipment>
        </clientRiskProfile>
</clientRiskProfiles>
```

This record shows the value 5 for **noClientRiskFactors**, meaning the client was asked about risk factors but none were identified. However, immediately below that, we see 1 for several specific risk factors, meaning that these risks were reported by the client, contradicting the value entered for **noClientRiskFactors**.

EvaluationWeb has validation checks in place to question this kind of scenario–when one variable relies on the response for another variable to make sense. Most often, you will receive warnings (as opposed to errors), guiding you to return to your data to verify whether the response values in the record are correct.

The screen shot below shows the DVS entry for the variable G216a (Vaginal or Anal Sex with a Male) and indicates where to find the information discussed above.

G216a   **Vaginal or Anal Sex with a Male**   **XSD (Schema) Name:**   **withMale**

Value Option: **Choose only one**   **Format Type:**   **Number**   **Min Length:** **1**   **Max Length:** **2**

Definition: The client self-reported having vaginal or anal sex with a male in the past 12 month

Instructions: Indicate if the client reported vaginal or anal sex in the past 12 months with a male.

Business rule   HIV Testing: Required, see detailed business rule
Partner Services: Required, see detailed business rule
HD Risk Reduction Activities: Required, see detailed business rule
CBO Risk Reduction Activities: Not applicable
HD Aggregate: Not applicable
CBO Aggregate: Not applicable

Detailed business rule:
Required if a client behavioral risk profile was collected (noClientRiskFactors=1)
Not expected if client was asked and no behavioral risk factors were reported, wasn't asked or declined to discuss behavioral risks (noClientRiskFactors= 5, 66 or 77).
Not expected if status of behavioral risk profile collection is missing (noClientRiskFactors = blank).

Look at the **detailed business rule** in the DVS for the variable.

Code	Value Description	Value Definition
0	No	*Client indicates that he or she did not have vaginal or anal sex with a male in the past 12 months.*
1	Yes	*Client indicates that he or she had vaginal or anal sex with a male in the past 12 months.*
99	Don't know	*Only select 'don't know' if the client states that he or she doesn't know if he or she had vaginal or anal sex with a male in the past 12 months. Do not select 'don't know' if the client was not asked.*

Additional things to remember

- Check the detailed business rules in the DVS for each variable.
- The values entered for some variables can influence whether other variables are required.
- The values entered for some variables can determine which value codes are acceptable for other variables.

(5) Data Set Specific Rules

Some validation rules only apply to certain data sets and not to others. Some examples are shown below:

An HIV Testing Data Set Example

```
...
<HivTests>
    <HivTest hivTestId="1">
            <sampleDate>01/02/2014</sampleDate>
            <workerId>14001</workerId>
            <testElection>1</testElection>
            <testTechnology>1</testTechnology>
            <testResultValueCode>1</testResultValueCode>
            <provisionOfResultValueCode>1</provisionOfResultValueCode>
            <reasonResultNotProvidedValueCode></reasonResultNotProvidedValueCode>
    </HivTest>
    <HivTest hivTestId="1">
            <sampleDate>01/02/2014</sampleDate>
            <workerId>14001</workerId>
            <testElection>1</testElection>
            <testTechnology>1</testTechnology>
            <testResultValueCode>1</testResultValueCode>
            <provisionOfResultValueCode>1</provisionOfResultValueCode>
            <reasonResultNotProvidedValueCode></reasonResultNotProvidedValueCode>
    </HivTest>
</HivTests>
...
```

In this example, **hivTestId="1"** has been used for both tests listed in the record. The FPR would generate the validation error message:

> "Cannot use the same Test Id twice within a form."

In a situation like this, you will not be able to find specific direction in the DVS since it concerns how the data in your file are organized instead of a DVS code or business rule violation.

The correct approach would be to list the HIV Test IDs sequentially, as seen below.

```
<HivTests>
  <HivTest hivTestId="1">
          <sampleDate>01/02/2014</sampleDate>
          <workerId>14001</workerId>
          <testElection>1</testElection>
          <testTechnology>1</testTechnology>
          <testResultValueCode>1</testResultValueCode>
          <provisionOfResultValueCode>1</provisionOfResultValueCode>
          <reasonResultNotProvidedValueCode></reasonResultNotProvidedValueCode>
  </HivTest>
  <HivTest hivTestId="2">
          <sampleDate>01/02/2014</sampleDate>
          <workerId>14001</workerId>
          <testElection>1</testElection>
          <testTechnology>1</testTechnology>
          <testResultValueCode>1</testResultValueCode>
          <provisionOfResultValueCode>1</provisionOfResultValueCode>
          <reasonResultNotProvidedValueCode></reasonResultNotProvidedValueCode>
  </HivTest>
</HivTests>
...
```

A Partner Services Data Set Example: Case Numbers

Before considering an example below, let's look at the nature of a Partner Services (PS) record.

In a PS record, the **localClientId** represents the client that the record was created for. The **localClientId** identifies a single client, who may be Index client, a Partner, or who may have both roles.

> ### Important!
>
> A Partner Services record for a single **localClientId** is not a listing of an Index client and his or her partners.
>
> Instead, it is a cumulative record containing various information about one client over time. Depending on the client's role(s), the record can include (among other things):
> - An Index client information section – this has the client's **partnerServiceCaseNumber** (if the client is an Index).
> - A Partner information section – this has the **partnerServiceCaseNumbers** of all the other cases that the client has ever participated in as a partner.

The information in the box above is important because it will help you avoid common errors in your Partner Services file uploads. It specifies that in a single record:

- There will be a **partnerServiceCaseNumber** which indicates the client's case number (if the client is an Index).

- There will other **partnerServiceCaseNumbers** which indicate the cases the client has participation in as a partner (if the client is a Parter).

- An Index client's **partnerServiceCaseNumber** is not time bound. On a practical level, this means:
  - One **partnerServiceCaseNumber** must follow the client throughout time.
  - Multiple sessions should be nested within the same case.
  - A case may be opened, closed, reopened, and closed again indefinitely.

Below is an example for the Partner Services version 2 data set. It contains a record for client 90899, whom we will call Jane.

Jane's **localClientId** is 90899. This ID identifies her as unique from all other clients in the system.

<clientlocalClientId="90899"clientid_status="N"clientid_lastModifiedDate="03/23/2015">

Jane is an Index client.

<indexPatientInformation>
<partnerServiceCaseNumber>123</partnerServiceCaseNumber>
...
</indexPatientInformation>

Jane's **partnerServiceCase Number** is 123.

<partnerInformation>
<partner>

Jane is a partner in another client's case. We will call this client John.

<partnerDateCollected>01/10/2015</partnerDateCollected>
<partnerServiceCaseNumber>456</partnerServiceCaseNumber>
...
<sessions>

John's **partnerServiceCase Number** is 456.

Jane has multiple sessions in a single **partnerInformation** node and in a single **partnerServiceCaseNumber** node.

<session>
<interventionId>227</interventionId>
<sessionDate>01/10/2015</sessionDate>
<siteId>227</siteId>
</session>
<session>
<interventionId>227</interventionId>
<sessionDate>03/12/2015</sessionDate>
<siteId>227</siteId>
</session>
</sessions>
</partner>
</partnerInformation>
</client>

Compare the above example with the one below.

> This is the same client as above, Jane, identified by **localClientId** 90899.

```
<client localClientId="90899" clientid_status="N" clientid_lastModifiedDate="03/23/2015">
```

> Jane's **partnerServiceCase Number** is 123.

```
    <indexPatientInformation>
        <partnerServiceCaseNumber>123</partnerServiceCaseNumber>
        …
    </indexPatientInformation>

    <partnerInformation>
        <partner>
```

> Jane is a partner in another case—again, John's.

```
            <partnerDateCollected>01/10/2015</partnerDateCollected>
            <partnerServiceCaseNumber>123</partnerServiceCaseNumber>
            …
            <sessions>
                <session>
                    <interventi
                    <sessionDa
                    <siteId>227</siteId>
                </session>
            </sessions>
        </partner>
```

> John's case number is indicated as 123. This is incorrect since 123 is Jane's case number. John's case number should appear here instead.

```
        <partner>
```

> Jane is a partner in yet another client's case. We will call this client Mary.

```
            <partnerDateCollected>03/12/2015</partnerDateCollected>
            <partnerServiceCaseNumber>123</partnerServiceCaseNumber>
            …
            <sessions>
                <session>
                    <interve
                    <session
                    <siteId>227</siteId>
                </session>
            </sessions>
        </partner>
    </partnerInformation>
</client>
```

> Mary's case is also indicated as 123. This is incorrect since 123 is Jane's case number. Mary's case number should appear here instead.

Since the **partnerServiceCaseNumber** (123) is repeated twice in the above record, EvaluationWeb would give the following validation error message in the File Process Report:

"Cannot use the same Test Id twice within a form."

The reason for the error is that when you repeat a **partnerServiceCaseNumber** in a single record, you are saying in essence that the client is an Index and he or she is a partner to him- or herself.

Things to remember

- Repeating a **partnerServiceCaseNumber** in a client record will generate an error, which will prevent you from uploading your file.
- You do not need separate **partnerServiceCaseNumbers** or separate partner nodes in a record in order to report multiple interviews (sessions). You can report a client's multiple sessions within a single partner node.
- Since XML is case sensitive, the XSD (schema) name should be formatted as **localClientId**, exactly as it appears in the schema.
- Use the Manage Agencies module in EvaluationWeb to check the setup information for your agency, program, intervention, and site(s), and update the setup accordingly.
    o For a Partner Services version 2 file, remember that sites are saved in a table separate from other agency information.
    o To check the existing PS sites setup in your instance of EvaluationWeb, see either the *PS v2 Sites* or *PS Sessions* Reflexx report.