Selecting a Schema for EvaluationWeb® File Uploads

Basic Information

What is a schema?
A schema is a set of rules for uploading data files to the EvaluationWeb® database. When you upload a file, it must match the rules specified in its corresponding schema.

What is the purpose of a schema?
The schema has 2 important purposes:

- It tells EvaluationWeb what kind of data you are uploading
- It tells EvaluationWeb how to check whether the data are organized in a consistent, logical way.

Who sets the schema rules?
Schema rules are determined by CDC in accordance with its data reporting requirements.

Are these the only rules the data are subject to?
No. Schema rules come into play only during the process of uploading a file. Once a file is uploaded, its data are subject to additional validation rules to ensure data quality. But before these additional rules can be applied, the file is first checked to make sure it matches the schema rules.

What happens when new schemas are released?
As data requirements evolve over time and as new data types emerge, new schema versions are released. In many cases, EvaluationWeb will accept older schemas, but eventually the older schemas become obsolete and EvaluationWeb stops accepting them. For example, the version 1.0 schema for HIV Testing data has been retired due to obsolescence.

What file formats and templates are available for upload files?
All data types may be uploaded in the XML file format, but some data types can also be uploaded in Microsoft Excel (xls or xlsx) or comma-separated values (CSV) formats. Excel and CSV files must also follow the schema rules, so Luther Consulting has provided flat file templates for these formats. See the next section for more information on file formats.

What data types can be key-entered into the database? What types must be uploaded?
In addition to being uploaded, HIV Testing, Risk Reduction Activities (RRA), and Agency Information data may also be key-entered into EvaluationWeb.

HIV Partner Services data must be uploaded.

Where can I find the schemas and templates?
All schema and template files can be found at: http://schemas.lutherconsulting.com

What schema versions are available for each data type? What file types are available?

The table below gives the schema versions and file formats that are available for each data type.

<table>
<tr><th>Data Type</th><th>Versions</th><th>File Types</th><th>Schema or Sample File Name</th></tr>
<tr><td>Agency Information</td><td>1.0</td><td>XML</td><td>AgencyInfo.xsd v1.0</td></tr>
<tr><td rowspan="3">HIV Testing</td><td>2.1
3.0
3.1</td><td>XML</td><td>CTData_v2_1.xsd
CTData_v3.0.xsd
CTData_v3.1.xsd (has revised housing status variable)</td></tr>
<tr><td>2.1
3.1
3.2</td><td>Excel*</td><td>CTRSample.xlsx (v2.1)
CTRTemplateV3.1.zip
CTRTemplateV3.2.zip</td></tr>
<tr><td>1.0 - retired</td><td></td><td></td></tr>
<tr><td rowspan="3">HIV Partner Services</td><td>2.0
2.1</td><td>XML</td><td>PSData_v2.xsd
PSData_v2.1.xsd</td></tr>
<tr><td>2.1</td><td>Excel*</td><td>PSv21Excel.zip</td></tr>
<tr><td>1.0 - retired</td><td></td><td></td></tr>
<tr><td>Client-level RRA</td><td>1.0 - HD**
1.0 - DF CBO***</td><td>XML</td><td>HDCLData_v1.xsd
CBOCLData_v1.xsd</td></tr>
<tr><td rowspan="2">Aggregate-level RRA</td><td>1.0 - HD
1.0 - DFCBO</td><td>Excel*</td><td>HDAGSample.xlsx
CBOAGSample.xlsx</td></tr>
<tr><td>1.0 - HD
1.0 - DFCBO</td><td>XML</td><td>HDAGData_v1.xsd
CBOAGData_v1.xsd</td></tr>
</table>

*.xlsx, .xls, or .csv
**HD = Health Department
†DF CBO = Community-based organization, directly funded by CDC

To better understand which schema you need, it is helpful to consider the way each schema handles the variables for its data type, as detailed in the tables on the next several pages.

Although the tables do not provide a comprehensive list of differences between the schemas, they give you the most important information needed to select the most appropriate schema to use for uploading your records.

Which HIV Testing schema is right for you?

	Version 2.1	Version 3.0	Version 3.1	Version 3.2
Schema (XSD) File	CTData_v2_1.xsd	CTData_v3.xsd	CTData_v3.1xsd	N/A
XML Sample File	HIVFormInfo_v2_1.xml	HIVFormInfo_v3.xml	HIVFormInfo_v3.1.xml	N/A
Excel Template File	CTRSample.xlsx	N/A	Add or Update Records: CTRSampleV3.1Tests.xlsx Delete Records: CTRSampleV3.1Delete.xlsx	Add or Update Records: CTRSampleV3.2Tests .xlsx Delete Records: CTRSampleV3.2Delete .xlsx
Agency Information	Builds out new sites if the Site ID in an HIV Testing record does not match information already saved in the EvaluationWeb database.	• Does not build out any new agencies, sites, interventions, or workers • All HIV Testing records must match agency, intervention, site, and worker ID numbers that are already in the EvaluationWeb database (these can be viewed in the Manage Agencies module) • Before uploading files, agency, site, intervention, and worker information must be entered via the Manage Agencies module or via an upload based on an Agency Info schema		

continued on next page

Which HIV Testing schema is right for you? (continued)

<table>
<tr><th></th><th>Version 2.1</th><th>Version 3.0</th><th>Version 3.1</th><th>Version 3.2</th></tr>
<tr><td rowspan="4">Housing Status Variable</td><td rowspan="4">Not reportable</td><td>Variable X730:
<housingStatus></td><td colspan="2">Variable X730a:
<housingStatusRevised></td></tr>
<tr><td>New as of the 04/16/2013 DVS</td><td colspan="2">New as of the 03/17/2014 DVS (replacement for variable X730)</td></tr>
<tr><td>Occurrences per record: 0 to 4 (“Select all that apply”)</td><td colspan="2">Occurrence per record: 1
(“Select only the most severe housing status”)</td></tr>
<tr><td>Valid values:
1 = Literally Homeless
2 = Imminently Losing Housing
3 = Unstably Housed and at-risk of Losing Housing
4 = Stably Housed
66 = Not Asked
77 = Declined to Answer
99 = Don’t Know</td><td colspan="2">Valid values:
1 = Literally Homeless
3 = Unstably Housed and/or at-risk of Losing Housing
4 = Stably Housed
66 = Not Asked
77 = Declined to Answer
99 = Don’t Know
(Imminently Losing Housing removed)</td></tr>
<tr><td>Client Zip Code</td><td>Not reportable</td><td colspan="2">Variable G134 <clientZipCode>
• Optional variable
• Can only be reported in XML; missing from Excel template</td><td>Variable G134 <clientZipCode>
• Optional variable
• Excel flat file template updated to include this variable</td></tr>
<tr><td>Local Use Fields</td><td>Not reportable</td><td>Not reportable</td><td>Not reportable</td><td>Available as optional columns
• localUseField1
• localUseField2
• localUseField3
• localUseField4</td></tr>
</table>

Which HIV Partner Services schema is right for you?

	Version 2.0	Version 2.1
Schema (XSD) File	PSData_v2.xsd	PSData_v2.1.xsd
XML Sample File	PSv2.0.xml	PSv2.1.xml
Excel Template File	Not applicable	PSv21Excel.zip
Index & Partner section names	<indexPatientInformation> <partnerInformation>	<asAnIndex> <asAPartner>
Collected Date for Client	Entered after each <client> node	Entered after the <asAnIndex> node
Attempt to Locate field set	Only one 'attempt to locate' can be present, only in the <indexPatientInformation> section	More than one 'attempt to locate' can be entered in the new <attemptsToLocate> node – can be entered in both the <asAnIndex> and the <asAPartner> sections
Elicit Partner field set	Only one partner elicitation may be present	More than one partner elicitation can be entered in the new <elicitPartners> node
Status	Attribute Z01: @status Valid values: A = Added Record D = Deleted Record N = New Record R = Resubmitted Record U = Updated Record	Attribute Z01: @status is not applicable since all added, new, resubmitted, or updated client records should be comprehensive anyway. Partner services sites will no longer be dependent on status Deleted records should appear in a separate <deletions> node with simplified information (agencyId and localClientId only)

There are no new variables between version 2.0 and 2.1. Instead, version 2.1 restructures existing data elements and removes some elements to simplify data submission. The elements that have been removed in version 2.1 are:

Variable A02 <populatedAreaValueCode>

Variable X503 <totalNumberOfPartnersClaimed>

Variable PCR202a <localPartnerServiceId>

Variable X303a <specifyReasonUnsuccessAttempt>

Which Client-Level RRA schema is right for you?

	Health Department Client Level Version 1.0	CDC Directly Funded CBO Client Level Version 1.0
Schema (XSD) File	HDCLData_v1.xsd	CBOCLData_v1.xsd
XML Sample File	HDCLSample.xml	CBOCLSample.xml
Program Announcement	Variable X137: <progAnnouncementProgStrategy>	Variable X137: <progAnnouncementProgStrategy>
	Occurrence per record: 1	Occurrence per record: 1
	Valid values: 1 = PS 12-1201 – Category A 2 = PS 12-1201 – Category B 3 = PS 12-1201 – Category C 7 = MSM Testing Initiative 10 = PS 12-1210 CAPUS 89 = Other (specify)	Valid values: 4 = PS 11-1113 5 = PS 10-1003 6 = PS 08-803 8 = PS 11-1113 Category A – YMSM 9 = PS 11-1113 Category B – YTG
	Note that variable X137 applies to both schemas, but the specific program announcement code will determine which schema to use.	

Which Aggregate-Level RRA schema is right for you?

	Health Department Aggregate-Level Version 1.0	CDC Directly Funded CBO Aggregate-Level Version 1.0
Schema (XSD) File	HDAgData_v1.xsd	CBOAGData_v1.xsd
XML Sample File	HDAGSample.xml	CBOAGSample.xml
Excel Template File	HDAGSample.xlsx	CBOAGSample.xlsx
Program Announcement	Variable X137: <progAnnouncementProgStrategy>	Variable X137: <progAnnouncementProgStrategy>
	Occurrence per record: 1	Occurrence per record: 1
	Valid values: 1 = PS 12-1201 – Category A 2 = PS 12-1201 – Category B 3 = PS 12-1201 – Category C 7 = MSM Testing Initiative 10 = PS 12-1210 CAPUS 89 = Other (specify)	Valid values: 4 = PS 11-1113 5 = PS 10-1003 6 = PS 08-803 8 = PS 11-1113 Category A – YMSM 9 = PS 11-1113 Category B – YTG
	Note that variable X137 applies to both schemas, but the specific program announcement code will determine which schema to use.	

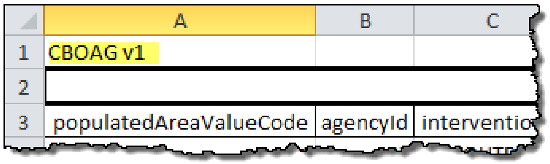
Which Agency Info schema is right for you?

	AgencyInfo Version 1.0
Schema (XSD) File	AgencyInfo.xsd
XML Sample File	AgencyInfo.xml

How do you declare which schema or you're using?

Part of using a schema is telling EvaluationWeb which schema your file is based on.

For Excel uploads, the flat file templates include the schema name and version number in cell A1, as seen here:



For XML files, it works differently:

- The XML specifies the particular kind of data you are uploading, as shown in the table below.

Data Type	How it should appear in the XML
HIV Testing	HivFormInfo
HIV Partner Services	xpemsPSData
Client-level RRA	clientData
Health Department Aggregate-level	HDAGData
Directly Funded CBO Aggregate-level data	CBOAGData
Agency Information	AgencyInfo

- The xsi:noNamespaceSchemaLocation value should point to the schema file itself. Even though the schema file name has the version number in it, the schema rules still require the version number to be specified separately from the file name. We recommend using the full web address of the schema file (an absolute link) rather than a local filename (a relative link) because the schema files located at http://schemas.lutherconsulting.com are always up-to-date and because it is an absolute link that CDC EvaluationWeb can always point to.

  <dataset xsi:noNamespaceSchemaLocation = "http://schemas.lutherconsulting.com/schemafilename.xsd" xmlns:xsi = "http://www.w3.org/2001/XMLSchema-instance">

  So, if you were uploading an XML file based on the HIV Partner Services version 2.0 schema, here's how you would indicate it at the beginning of your file.

  <xpemsPSData xsi:noNamespaceSchemaLocation = "http://schemas.lutherconsulting.com/PSData_v2.xsd" xmlns:xsi = "http://www.w3.org/2001/XMLSchema-instance">

Likewise, if you were uploading an XML file based on the CBO Aggregate-level RRA version 1.0 schema, you would indicate it like this:

<CBOAGData xsi:noNamespaceSchemaLocation = "http://schemas.lutherconsulting.com/CBOAGData_v1.xsd" xmlns:xsi = "http://www.w3.org/2001/XMLSchema-instance">

> ***TIP***
>
> All of the sample files show the correct way to display this information for a given schema

---

Questions?

This document is an introduction to some common considerations for selecting the appropriate schema or template for uploading your data. 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