HMIS XML FY2022 Data Exchange Format Specifications v1.0.0

aligns with HMIS Data Standards Version FY2022, effective October 1, 2021

HUD HMIS XML Schema Documentation

9/22/2021, version: FY2022v1.0.0

This document was developed for the US Department of Housing and Urban Development (HUD), Office of Special Needs Assistance Programs (SNAPS) in the Office of Community Planning and Development, under a contract with Abt Associates, Inc..

Primary author:
Eric Jahn, Alexandria Consulting LLC

Table of Contents

Document changes with this version are available in the online versioning system for version FY2022.

Overview

Audience for this Document

This document is intended for developers and project managers implementing software systems which need to transmit HUD HMIS project and client information over a network. Because of this, it is inherently a technical document. Those new to eXtensible Markup Language (XML) should first familiarize themselves with XML technology in general, before examining the actual schema product and the "Technology Used" section of this document.

Purpose of this Document

The US Department of Housing and Urban Development (HUD), in cooperation with the Department of Health and Human Services (HHS) and the Department of Veterans Affairs (VA), is responsible for the Homeless Management Information System (HMIS) Data Standards, which define data collection requirements for any software used as an HMIS. The first version of the data standards was published in 2004; the current version (FY2022) has an effective date of October 1, 2021.

The HUD HMIS XML Schema, version FY2022 specifies a format for transferring this HMIS data. The data elements are defined in the FY2022 HUD HMIS Data Manual, and further specified by the FY2022 HUD HMIS Data Dictionary and FY2022 HMIS Logical Model. The HUD HMIS XML Schema should also be consistent semantically and logically with the HUD HMIS CSV, though more normalized. The Logical Model, Manual, and Dictionary all changed between versions FY2020 and FY2022, so the XML Schema has been changed accordingly. The key benefit of XML is its ability to be validated and extended. The XML Schema also comprises a fully machine readable specification, unlike for CSV, where the data types must be manually read from a document. XML Schema's validation capability means that systems, both those that send and receive data, can independently and automatically check if this data is being sent to the HMIS XML specification. This helps to evaluate data quality and correctness independently, without either transmitting party (sender/receiver) needing to develop custom validation software. The HMIS XML Schema can be custom extended, as well, but still keep the ability to check the generated data for accuracy against this original HMIS XML Schema. The resulting valid XML format can be used for data migrations between systems, or the data types defined within the schema can be individually referenced in custom web APIs.

Alternative

For HMIS implementers seeking an alternative format for HMIS client data transmission, see the equivalent HUD Comma-Separated Values (CSV) for HMIS data, version FY2022[1]. Both the XML and CSV sister formats are based on the same HMIS Logical Model, and are completely semantically and logically compatible because of this fact. Differences in element groupings exist, due to file count considerations for the CSV, which is denormalized from the Logical Model.

Use Cases and Extension

There are a number of purposes for which HMIS data might be exported from one system and imported to another. The use cases that the HMIS CSV and HMIS XML formats are primarily intended to support include migration from one HMIS application to another; warehousing of data from multiple HMIS implementations for analysis and reporting; and participation in a local HMIS implementation by regularly providing data entered

[1]see https://www.hudhdx.info/VendorResources.aspx

into and exported from an alternate database. In general, HUD expects that it should be possible to export, in a standard format, all data entered into an HMIS for any data element defined by the HMIS Data Dictionary, regardless of whether or not a given data element is required based on project type or funder. Specific requirements related to the overall functionality of HMIS applications and export and import processes in particular will be established by the forthcoming HMIS Software and Data Quality Standards and may result in revisions to this document. The HMIS Data Dictionary defines several hundred fields. HUD is aware that there is, in addition, a wide variety of expanded and customized data collection in HMIS implementations across the country. While HUD is cognizant that the exchange of additional data that may be included in an HMIS will often be useful and necessary, it is not practical to include accommodations for every potential need, even if it were possible to anticipate them. As a result, the scope of this document is generally limited to data collected in a manner consistent with the HMIS Data Dictionary. The HMIS XML format may be extended to include additional fields by parties engaged in HMIS data exchange; technical assistance may be available. To request technical assistance please visit www.hud ehttps://www.hudhdx.info/Resources/Vendors/HMIS_CSV_Specifications_FY2022_v1.0.pdfxchange.info.

Scope of the HMIS XML Schema

This schema requires complete data sets for each grouping of data (groups like Inventory, Project, Path Status, etc.), to accommodate the use case of transmitting complete datasets for reporting, periodic bulk data uploads from a partner agency to an HMIS or data warehouse, or for HMIS system migration purposes. Because of this intended use case, many of the data elements are not optional. For example, if an XML document following the Schema includes any of the data elements within the "Funder" class of the Logical Model, it must include almost all of the required data elements within Funder (since for Funder, all its elements are required, except one). In past versions of the HMIS XML Schema (prior to v4.0), most data elements were listed as optional, to accommodate use cases other than reporting complete data sets. For other use cases, the individual data groups could be individually transmitted in bulk (by removing the key and cardinality restrictions/assertions within the Export element), but that is beyond the scope of this current specification. It should be possible to export, in a standard format, all data entered into an HMIS for any data element defined by the HMIS Data Dictionary, regardless of whether or not a given data element is required based on project type or funder.

Location of the HMIS XML Schema

The HMIS XML Schema version FY2020 and previous major releases are hosted at http://www.hudhdx.info/VendorResources.aspx. All releases are also kept in a versioning system.

Components

This publication has multiple artifacts, and each artifact is available in two places: the HUD HDX link above, and a development website, where new versions are discussed, issues logged, and all incremental changes are stored. Each artifact is enumerated below.

1. This document, which includes:
    - A rationale for the schema, including an overview of the process, and an explanation of the model.
    - A description of the steps involved beyond creation of a data standard, including development of communication protocols and documentation of responsibilities.
    - A brief discussion of the future path of HMIS XSD development.

2. An XML Schema Definition (XSD) document: HMIS XSD version FY2022. It contains inline documentation which correlates each schema element to an item in the HMIS Data Dictionary and Data

Manual. Searching the HMIS XML Schema for the corresponding HMIS Data Dictionary data element numeral provides a cross-walk between technical schema and the HMIS Data Standards.

3. A sample, valid XML document with fictitious data. A minimal instance is also available there for testing.
4. An example extension schema of the HMIS XSD version FY2022, illustrating how to add an additional data element.
5. A sample, valid XML instance document for the extended HMIS XSD version FY2022.
6. Online, browsable, graphical documentation for HMIS XSD version FY2022.

Overall Structure

The HMIS XML version FY2022 structure is intended to be completely compliant with its parent specification, the FY2022 HUD HMIS Logical Model (released July 2019), and with its other similarly purposed product, the HUD HMIS CSV version FY2022. To keep its structure simple, the XML continues to be composed of a relatively flat, single file. Almost all the data types are one level below the Export data element and key references enforce relationships between types. This allows the system serializing the XML to not be constrained by having to be in the correct context to add elements. Elements can be added in any order within the "Export" element.

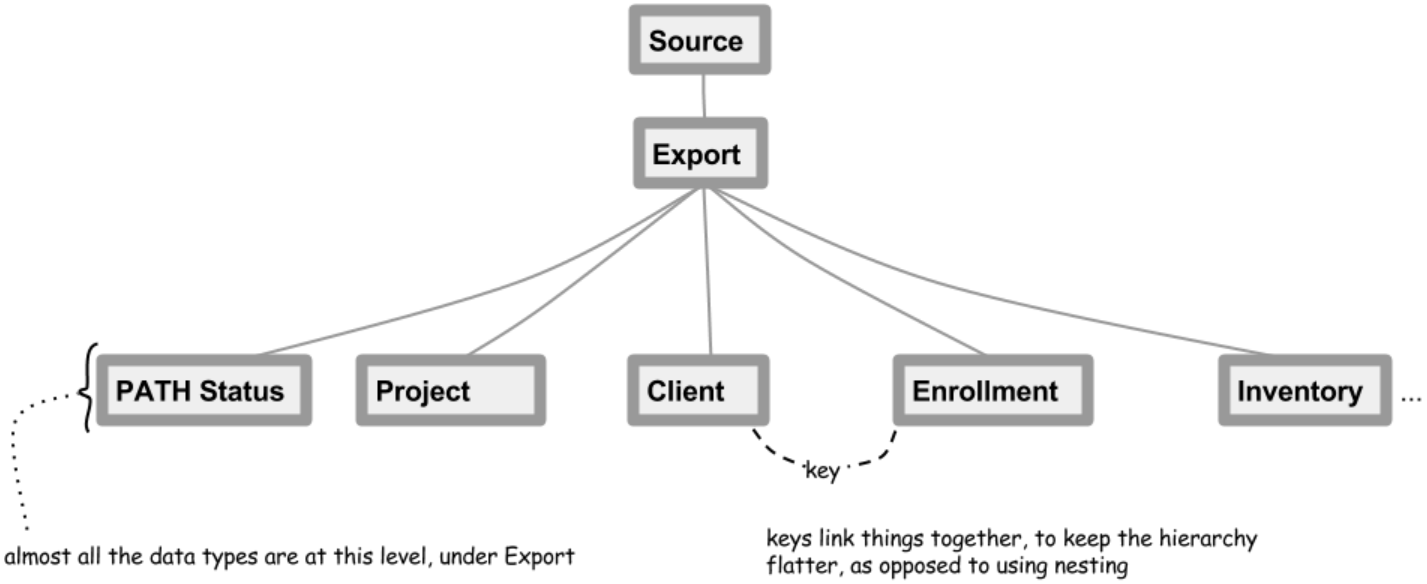
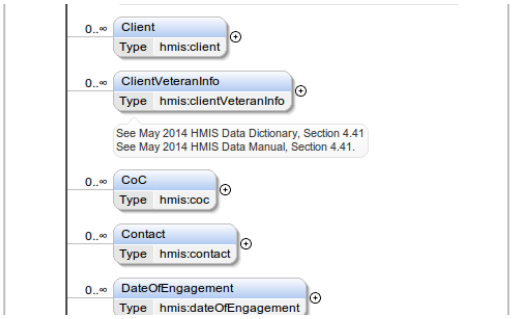


Figure 1: HMIS XML Basic Structure

Inline Documentation

The HMIS Data Dictionary and Data Manual data element reference is listed within each HMIS XML data element, whenever available. Some data elements, such as Export, do not reference a data element in the Dictionary/Manual, since they only serve a functional purpose for data transfer.

Browsable Graphical Documentation



A browsable list of all the data elements in the HMIS XML Schema is available. The definitions for each data element are available in the browsable documentation as well.

Relationship to Logical Model

The HMIS XML Schema version FY2022 is conformant with the HMIS Logical Model, The various relationship lines within the Logical Model are enforced by requiring matching IDs/Keys (see the section on Keys) between the two data types (like Project to Inventory, or Exit to Enrollment) joined by a line in the Logical Model.

Cardinality Enforcement

The Logical Model is also very specific about cardinalities (like “many-to-many” and “zero to one”) between the related data types. HMIS XML Schema strictly enforces cardinality by means of XPath 2.0 assertions[2]; a new feature of XML Schema 1.1. If the Logical Model requires “zero to one” of something (like Date of Engagements per Enrollment), the HMIS XML version FY2022 will likewise require valid XML documents to declare the required zero or one occurrence.

Changes from Previous Version

All data elements are updated from the HUD HMIS Data Standard 2020 version 1.4 (released in August 2019) to the HUD HMIS Data Standard FY2022 version 1.0 (released May 2021) elements.
Line-by-line comparison of version FY2020 to version FY2022:
https://github.com/hmis-interop/xml/compare/FY2020-latest...FY2022-latest (...and click ”Load diff”. This will stay updated with later bug-fixes). For a less concise list of changes, see the Overview of Changes below. For new elements, conditional validation has been added for many of the newly added dependent fields, where specified in the Data Dictionary. See the section below on Conditional Validation.

Technology Overview

XML Schema 1.1

XML Schema version 1.1 is the format and standard used by the HMIS XML Schema version FY2022. It was first introduced with HMIS XML Schema 4.0. XML Schema 1.1 is a superset of the XML Schema 1.0, used in version 3.1 and below of the HMIS XML format, and is forward compatible from XML Schema 1.0. XML Schema 1.1 only adds additional features to those of 1.0, three of which are used by HMIS XML Schema FY2022. These are extensibility features, conditional validation, and better key reference enforcement features. All major XML parsers have been updated to support version Schema 1.1, which was released in

[2] http://www.w3.org/TR/xmlschema11-1/#cAssertions

2007, resulting in XML Schema 1.1 now being a mature technology. A practical example of command line validation using the Xerces XML Parser is in the wiki.

Extensibility

Flexible vendor extensions are much more easily accomplished with XML Schema 1.1, as opposed to the previous XML Schema 1.0. To extend the HMIS XML Schema version FY2020 using this new feature, please read the documentation on *defaultOpenContent mode="interleave"*, *xs:redefine*, and *xs:override* and at the The World Wide Web Consortium (W3C) website. This new feature permits is implemented in HMIS XML, allowing the addition of new data elements anywhere in the HMIS XML Schema, and the original parts still can be found and validated. The newly added extensions which a software provider might add should be validated by a second XML Schema. The second XML Schema can be defined elsewhere, and both the HMIS Schema and the new extension schema can function in concert to validate the extended XML documents. The extended schema will still validate against the official HUD HMIS XML Schema, but it will not invalidate the custom elements added.

Currently, HUD HMIS is not publishing any official extensions, but the parties engaged in data integration may negotiate the use of an extension schema to be overlaid onto the HUD HMIS XML Schema for the addition of new data types, attributes, and elements.

Hashing

Hashing is possible for personal identifiers. There are three possible values for the “hashStatus” attribute:

- “Unhashed” - Data are expected to conform to data type and field length parameters defined in the schema.
- “SHA-1 RHY” - Note: For federal data exchange purposes, this is a legacy methodology. RHY requires use of SHA-256 (see below). The SHA-1 algorithm produces a 40-character string of letters and numbers. More specifically, for the following identifiers, a specific hashing method is required:
    - FirstName – SHA-1 hash of the SOUNDEX of the value for first name;
    - MiddleName (S40) – SHA-1 hash of the SOUNDEX of the value for middle name;
    - LastName (S40) – SHA-1 hash of the SOUNDEX of the value for last name; and
    - SSN (S44) – concatenation of the unhashed last 4 digits of the SSN followed by SHA-1 hash of the full SSN.
- “Other” - Identifies exports in which data are hashed based on parameters mutually agreed upon by the sender and recipient of the XML data.
- “SHA-256” - The SHA-256 algorithm produces a 64-character string of letters and numbers. Data sets with a HashStatus of ‘SHA-256’ (4) will be exported consistent with the specifications of this document with the exception that the following fields in Client element (and no others) will be hashed using the SHA-256 standard algorithm and data types for these four fields will differ as noted: • FirstName (S64)– SHA-256 hash of the SOUNDEX of the value for first name; • MiddleName (S64) – SHA-256 hash of the SOUNDEX of the value for middle name; • LastName (S64) – SHA-256 hash of the SOUNDEX of the value for last name; and • SSN (S68) – concatenation of the unhashed last 4 digits of the SSN followed by SHA-256 hash of the full SSN. In the case of a partial SSN, use a lower-case letter x to replace any missing digits. The resultant 9-character string will be hashed in the same manner as a complete SSN.

Keys

The HMIS XML schema uses keys and references to those keys (aka “keyrefs”) to create the relationships defined in the Logical Model. The keys are enforced, so that if an enrollment in the XML references a project ID which isn’t defined somewhere else in the XML, it will raise a validation error, which enforces complete,

self-referential data sets. Keys also enable the flattened structure of the schema. An advantage of a flattened schema is that data elements are not repeated within deeply nested structures. Keys also makes programming simpler, since there are not so many nested logic structures to handle. The flattened structure is also more flexible, since not every related data element has to be mentioned within the same tree branch of the XML.

Conditional Validation

The ability to check for a specific precondition before validating is employed in HMIS XML Schema version FY2022. The main use of conditional validation in the HMIS XML Schema is to first check whether a previous question has been answered in a specific way, before allowing a follow-up question to be asked. As with the CSV specification, dependent fields will be null unless the parent field contains a value defined by the HMIS Data Dictionary as triggering the dependent field. Conditional validation is only applied for most newly added questions introduced in 2022 and prior versions of the HUD HMIS Data Standard, so the coverage of all possible conditional logic is not comprehensive. The main purpose of this conditional validation is to catch typical problems, and to enforce some of the most clearly described question workflows in the Data Standard. Conditional logic is implemented using XPath 2.0 assertions, similar to the existing key reference and cardinality enforcement assertions. This conditional validation can slow the validation time for large XML files, so it can be turned off by simply commenting out the various lines in the version FY2022 Schema beginning with “assert”. A regular expression run to remove or comment out all lines in a schema that begin with <xs:assert> would accomplish this task, and allow for “primary validation” which excludes the more time consuming “secondary validation” with conditional assertions. A good testing approach is to test small development XML instance files first with secondary conditional assertion validation, and then use the primary validation for fast production validation of large files.

Here is an example of conditional validation in the schema (at line 546), using an assertion:

```
<!--Testing if DisabilityType is 8, before allowing use of TcellCountAvailable-→
<xs:assert test="if (hmis:TCellCountAvailable) then boolean(hmis:DisabilityType/text()='8') else
true()"/>
```

Examples

Example Instances

A fairly exhaustive example HMIS XML instance is available in the repository for download. A minimal instance is also in the repository for download.

Example Custom Extension and Instance

An example HMIS XML extension is available in the repository for download. An instance applying this extension is also in the repository for download.

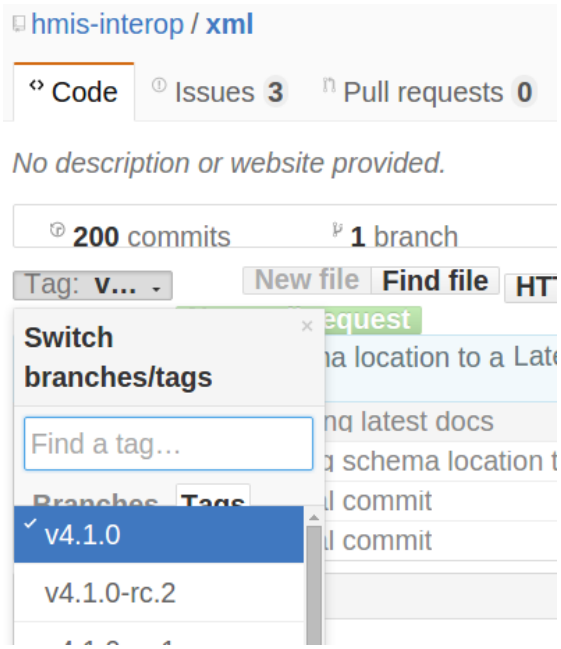
Change Process

To request changes to the HUD HMIS XML Schema, there are multiple ways to register requests.

- A software issue tracker is available at: https://github.com/hmis-interop/xml/issues
- A HUD sponsored HMIS vendor forum has monthly meetings, where the HMIS Data Standards are often discussed. Use the HUD Ask-A-Question (AAQ) link below to request a login to participate in the calls.
- Contact HUD AAQ to make a request to HUD HMIS Technical Assistance

HMIS XML Schema Hosting

HMIS XML Schema development archives and issues list is unofficially (not an official HUD website) hosted at: https://github.com/hmis-interop/xml. The tags on that site contain the various versions available, from which the differences between versions can be listed in detail.



Further Information

Contact the HUD Exchange's "Ask A Question" portal to request general assistance with the HMIS XML Schema. To log a detailed change directly, an issue tracker is available.

Overview of Changes

Change List

1. FY2022 HMIS XML Schema changes from FY2020 – all changes listed here: https://github.com/hmis-interop/xml/compare/FY2020-latest...FY2022-latest (click the "Load diffs" link to expand).

2. All 2022 HMIS Data Dictionary element and picklist changes, matching the changes in the FY2022 HMIS CSV

3. The only significant structural change is the addition of machine readable unique official HMIS Data Dictionary data element IDs, and verbatim machine readable Data Dictionary question text. This could be useful for automatically generating system tables containing HMIS questions and IDs.

   For example in the 2022 XML Schema:

```
<xs:element name="AssessmentLocation" id="_4.19.2" type="hmis:string250">
        <xs:annotation>
                <xs:documentation xml:lang="en">Assessment location</xs:documentation>
        </xs:annotation>
</xs:element>
```

the documentation (minus the bracketed format note) is copied verbatim from this Data Dictionary instruction listing:

4.19 Coordinated Entry Assessment

Header	Instruction
Element Name	Coordinated Entry Assessment
Field 1 & Response	Date of assessment [date]
Field 2 & Response	Assessment location [drop down]

There is only one element with a given ID attribute in the HMIS XML Schema, so there will never be duplicate IDs in the XML schema.

4. Disabilities and Financial Assistance subtypes are reworked so each subtype has its own separate element definition in the schema with the picklists formally enumerated in the schema, so the picklist values are machine readable. This also allows each subtype to have its proper machine readable FY2022 Data Dictionary numbering.