caBIG cancer Biomedical Informatics Grid

caAdapter

Use Case: Mapping Tool

Version 1.3

NATIONAL CANCER INSTITUTE®

Center for Bioinformatics

DOCUMENT REVISION HISTORY

Version Number	Date	Description
1.2/1.3	April 1, 2005	• First Draft
1.2/1.3	April 14, 2005	• Second Draft
1.2/1.3	July 29, 2005	• Edited Second Draft
1.2/1.3	August 17, 2005	• Minor edits to v1.2 per document specialist review
1.2/1.3	Feb 6st, 2006	• Rename to caAdapter, add H3S use case

Table of Content

1 Background/Summary

The caAdapter supports NCICB's mission of developing a translational research infrastructure and building a clinical research network by providing a common platform for sharing data. The caAdapter currently facilitates the building, parsing and validation of Health Level Seven Version 3 (HL7 v3) messages based on specific schema definitions. The mapping tool provides users a graphical user interface (GUI) and a transformation service to map and convert clinical data in a comma separated value (CSV) format to an equivalent target HL7 v3 XML format.

2 Introduction

2.1 Overview

The caAdapter Mapping Tool is developed as an open source application that enables analysts and developers to create a mapping from one data format to another. The first phase of mapping tool development focuses on mapping from a CSV format to an HL7 v3 message structure. Other incoming and outgoing formats, possibly including non-HL7 XML, HL7 v2 messages, database and Java objects, may be considered in later phases. Specifically, this mapping tool is designed as an easy to use GUI for defining source and target specifications, mapping source to target, using the map to convert actual data files and validating each step in the process.

2.2 Related Documents

caAdapter Phase 2 Scope

Requirements Specification for caAdapter Mapping Tool

3 Actors and Goals

3.1 HL7 Clinical Data Analyst

An HL7 clinical data analyst is a user that is expected to understand basic concepts about HL7 and the particular target message structure as well as the content and organization of the clinical data source file. This user is the subject matter expert whose goal is to perform the item-by-item mapping of source data to target message elements and to generate a mapping file to use in a data conversion process.

3.2 External System

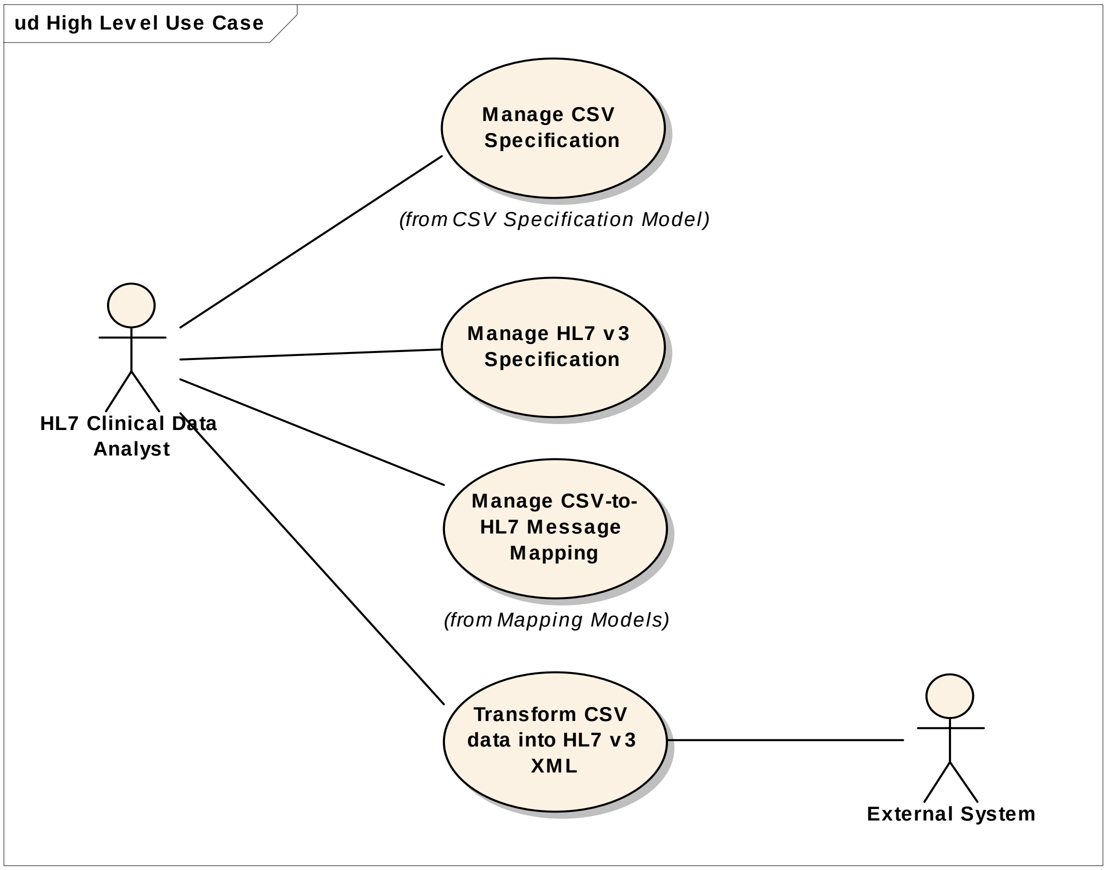
An external system may be any one or more of number of consumers within the clinical trials architecture. Such consumers may include message exchange services, transaction databases, research databases, or even external organizations and entities.

4 Use Cases

This section begins with four use cases related to the conversion of incoming CSV files. The first use case is *High Level Use Case*. The following three use cases, CSV Specification Use Case, HL7 v3 Specification Use Case, and Mapping Specification Use Case, expand into the process of defining CSV specification files and HL7 v3 specification files, and creating mappings.

4.1 High Level Use Case

4.1.1 Use Case Diagram


ud High Level Use Case
Manage CSV Specification
(from CSV Specification Model)
Manage HL7 v3 Specification
Manage CSV-to-HL7 Message Mapping
(from Mapping Models)
Transform CSV data into HL7 v3 XML
HL7 Clinical Data Analyst
External System

4.1.2 Use Case ID: caAdapter_Mapping-UC-01

4.1.3 Brief Description

In this scenario, the HL7 Clinical Data Analyst has already, or plans to generate, a CSV file in the form of logical records comprised of one or more lines (segments) each beginning with an identifier (segment name) which is followed by one or more data items (fields). The Mapping Tool allows the analyst to create and edit a CSV specification describing the file format of the CSV file, create and edit an HL7 v3 message specification, and then map from a CSV specification to an HL7 v3 message specification. The resulting mapping file can then be used by the message exchange service to convert CSV files into standard HL7 v3 XML messages.

4.1.4 Primary Actor

HL7 Clinical Data Analyst

4.1.5 Preconditions

4.1.5.1 Existing CSV File

To begin this process, the HL7 Clinical Data Analyst has already, or plans to generate, a CSV file from a clinical system conformance with the following prescribed characteristics.

- File contents are organized into multi-line logical records.
- Each line, called a segment, begins with an identifier, called a segment name, and is terminated by a new-line character.
- Each segment has one or more data items, called fields, which follow the segment name and terminates by commas (except for the last field on the line that uses the segment terminator).
- Segments may occur more than once in the same logical record, except for the first or root segment, which always indicates the beginning of a new record.
- Segments are related to one another in a parent-child hierarchy that documents the one-to-many nature of the association between related data items.
- A CSV file may have one or more logical records. Each of these is terminated by the beginning of the next record (a new root segment) or the end of file.
- The intention is that each logical record will generally become one single HL7 v3 XML message instance.

4.1.5.2 HMD File

To complete this process, the analyst must have an HL7 Hierarchical Message Description (HMD) corresponding to the type of data in the CSV file from the source system. This HMD describes the target message type that will be generated from the CSV data.

4.1.6 Basic Flow of Events

4.1.6.1 Actor, the analyst, using the mapping tool GUI, selects one of four methods (documented in detail in CSV Specification Use Case) to generate a validated CSV specification that describes the structure of a CSV file.

4.1.6.2 Actor, the analyst, using the mapping tool GUI, generates (documented in detail in HL7 v3 Specification Use Case) a validated HL7 v3 specification for the target message type.

4.1.6.3 Actor, the analyst, selects a CSV specification as the source and an HL7 v3 specification as the target, and the GUI displays them in a tree-like hierarchy on the left and right sides of the screen. The analyst uses a drag-n-drop process (documented in detail in the Mapping Specification Use Case) to create a validated mapping file between the source and target fields.

4.1.6.4 Actor, the analyst, selects the CSV data file and the mapping file, and converts CSV data into a validated HL7 v3 XML instances.

4.1.7 Post Conditions

4.1.7.1 The system successfully generates HL7 v3 XML instances without any warning or error.

4.1.7.2 The system generates HL7 v3 XML instances with warnings.

4.1.7.3 The system generates HL7 v3 XML instances with errors.

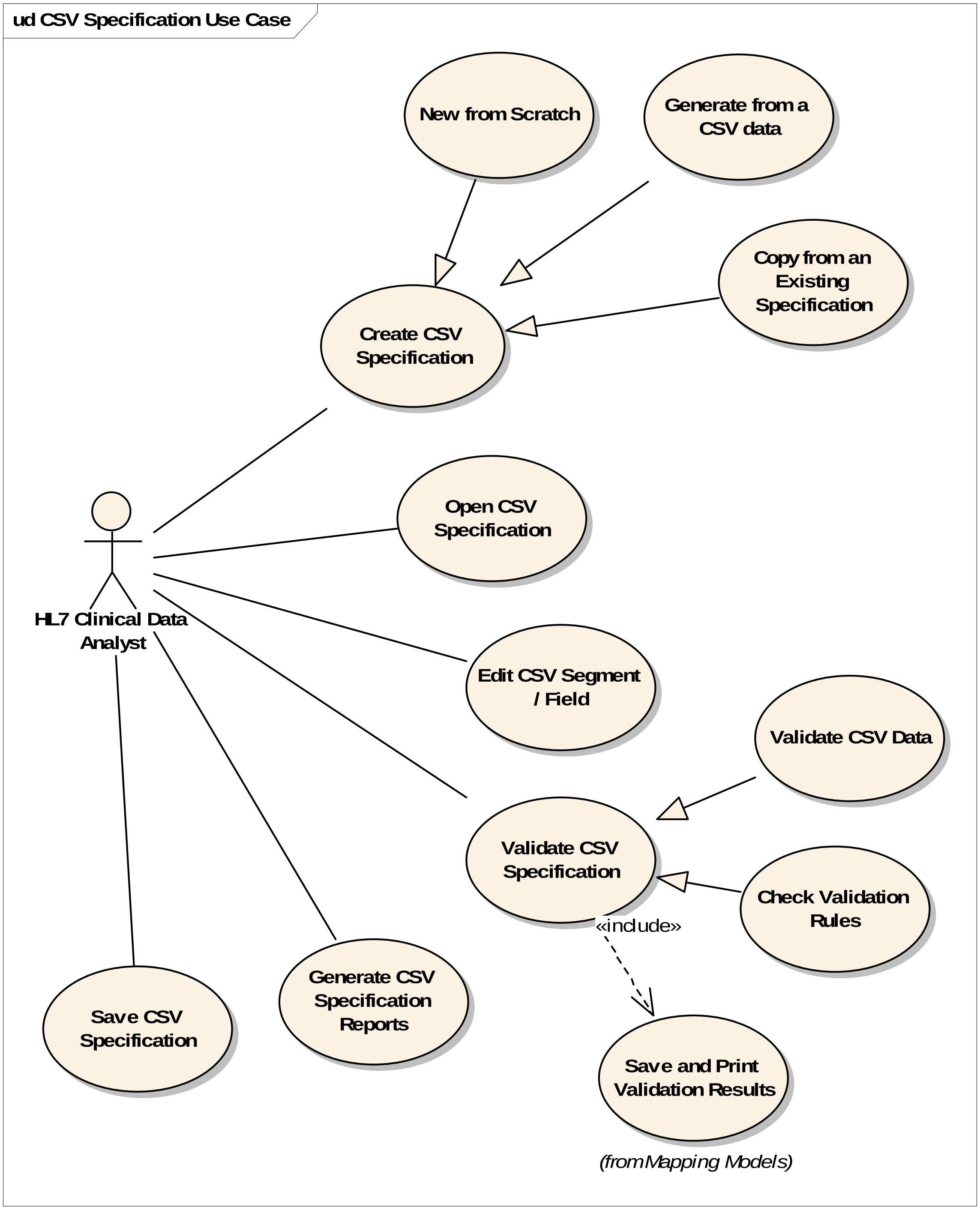
4.1.7.4 The system fails to generate HL7 v3 XML instances because of a fatal error.

4.1.8 Alternative Courses of Action

4.1.8.1 The system fails to generate HL7 v3 XML instances or generate HL7 v3 XML instances with warnings and/or errors, and then the analyst continues modifying the mapping file (returns to step 4.1.6.3).

4.2 CSV Specification Use Case

4.2.1 Use Case Diagram



4.2.2 Use Case ID: caAdapter_Mapping-UC-02

4.2.3 Brief Description

In this scenario, the HL7 Clinical Data Analyst has or plans to generate a CSV file in the form of logical records comprised of one or more lines (segments) each beginning with an identifier (segment name) that is followed by one or more data items (fields). The Mapping Tool allows the analyst to initialize the CSV specification that documents such a file structure using one of four methods. A CSV specification file may be:

a. created based on an existing CSV file
b. created from scratch
c. created based on an existing CSV specification file or
d. opened from an existing one

Once the specification has been initialized, the Mapping Tool allows the user to edit it, validate it, save it to a file and thus share it with other analysts. The resulting CSV specification file is used in the Mapping Specification Use Case.

4.2.4 Primary Actor

HL7 Clinical Data Analyst

4.2.5 Precondition

4.2.5.1 Existing CSV File

a. The HL7 Clinical Data Analyst has already, or plans to generate, a CSV file from a clinical system.

4.2.6 Basic Flow of Events

4.2.6.1 Create or Open CSV Specification

a. New From a CSV Data File - The analyst selects an existing CSV file from the file system.

   - Parse the file to identify distinct segment names, including the first or root segment, and the maximum number of fields present for each of the segments individually.

   - Display the list of segments and their fields in a tree-like outline. All segments are initially children of the root segment, and all fields are initially called “field1”, “field2”, etc.

b. Create From Scratch - The analyst starts a new source tree hierarchy. This hierarchy has an initial root segment with a default segment name and no fields.

c. Copy from an Existing CSV Specification.

d. File > Open - The analyst selects an existing CSV specification file from the file system using a Windows File > Open type of option in the Mapping Tool GUI. The system reads in the CSV specification and displays it to the user in the GUI. This data is editable and can be saved as the same original file or as a new file. The analyst reviews the list of segments and fields, rearranges the parent-child associations between segments (with the first segment remaining the root), adds new segments and fields, and renames or deletes unwanted segments and fields until satisfied with the hierarchy.

4.2.6.2 The analyst validates the CSV specification:

a. Validate CSV data against the specification.

b. Validate the CSV specification against validation rules.

c. The analyst saves and/or prints the validation results.

4.2.6.3 The analyst may generate a CSV specification report.

4.2.6.4 The analyst saves (or saves as) the CSV specification to file in an XML format that can be opened by another analyst who can continue working on the specification.

4.2.7 Post Conditions

4.2.7.1 The system successfully manages the CSV specification without any validation warning or error.

4.2.7.2 The system manages CSV specification instances with warnings or errors.

4.2.7.3 The system fails to create or open the CSV specification because of a fatal error.

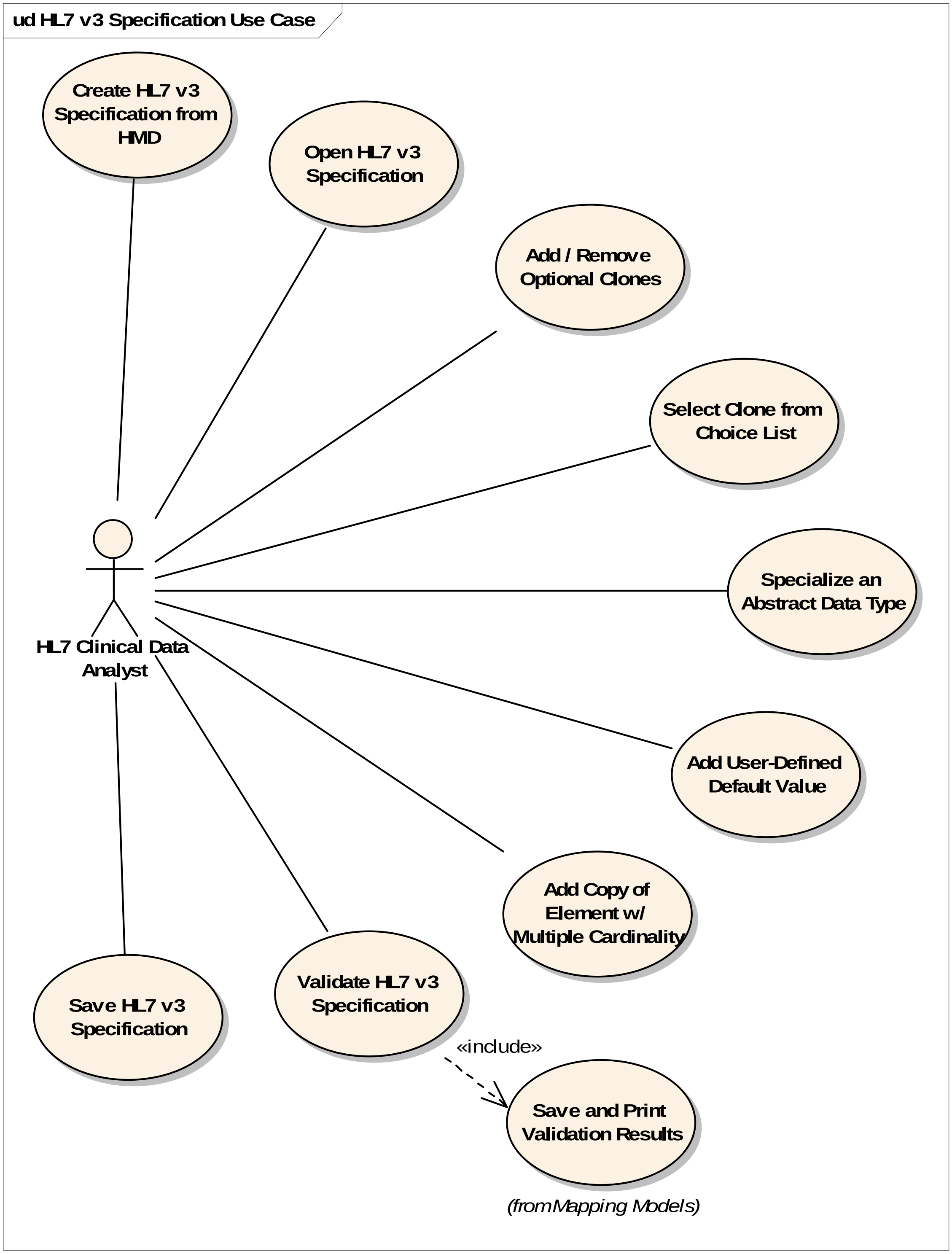
4.2.8 Alternative Courses of Action

4.2.8.1 The system manages the CSV specification with warnings and/or errors, and then the analyst will modify the CSV specification (returns to step d).

4.2.8.2 The system fails to create or open the CSV specification. The actor reports the error.

4.3 HL7 v3 Specification Use Case

4.3.1 Use Case Diagram


ud HL7 v3 Specification Use Case
Create HL7 v3 Specification from HMD
Open HL7 v3 Specification
Add / Remove Optional Clones
Select Clone from Choice List
Specialize an Abstract Data Type
Add User-Defined Default Value
Add Copy of Element w/ Multiple Cardinality
HL7 Clinical Data Analyst
Save HL7 v3 Specification
Validate HL7 v3 Specification
«include»
Save and Print Validation Results
(fromMapping Models)

4.3.2 Use Case ID: caAdapter_Mapping-UC-03

4.3.3 Brief Description

In this scenario, the HL7 Clinical Data Analyst has identified an HL7 v3 message type to which the CSV data must be converted. Because an HL7 v3 HMD may have many ambiguities which make mapping and data conversion difficult, the analyst must create a specification which tailors the message type to one particular purpose. The specification is an XML document identifying message elements found in the HMD and additional elements including fields for each attributes data type and user-defined default values. The Mapping Tool allows the analyst to create and edit this HL7 v3 specification using one of two methods. An HL7 v3 specification file may be:

a. based on an HL7 HMD or
b. based on an existing HL7 v3 specification file.

Once the specification has been created, the Mapping Tool allows the user to edit it, save it to a file and thus share it with other analysts. The resulting HL7 v3 specification file is used in the Mapping Specification Use Case on page 18.

4.3.4 Primary Actor

HL7 Clinical Data Analyst

4.3.5 Secondary Actor

Second HL7 Clinical Data Analyst (inherited from the main *High Level Use* Case on page 7 though not explicitly displayed in this diagram)

4.3.6 Precondition

4.3.6.1 Existing HMD File

a. To begin this process, the HL7 Clinical Data Analyst has already integrated into the tool the HMD for the desired HL7 v3 message type. (Instructions for such integration are found at User Guide.)

4.3.7 Basic Flow of Events

4.3.7.1 Create or Open CSV Specification

a. Create New Based on an HMD - The analyst selects an Message Type from a dropdown list.

- Parse the HMD to identify not just clones and attributes, but also to extend structure to include the fields associated with the attributes' data types (a.k.a. data type fields).

- Display the list of clones, attributes and data type fields in a tree-like outline, in a hierarchy consistent with the HMD.

b. Open an Existing HL7 v3 Specification - The analyst selects an existing HL7 v3 specification file from the file system using a Windows File > Open type of option in the Mapping Tool GUI. The system reads in the specification and displays it to the user in the GUI. This data is editable and can be saved as the same original file or as a new file.

4.3.7.2 Adjust the Specification to Suit

a. Add or Remove Optional Clones - Clones that are not mandatory or required or that have a minimum cardinality of 0 are effectively "optional" and don't necessarily have to be in the specification unless they are needed to carry source data. The analyst adds or removes such clones along with their substructure (associated clones, attributes and data type fields).

b. Select Clones From Choice List - For HL7 messages that include a choice box, the Mapping Tool presents the analyst with the option to choose one clones from the list of those in the box.

c. Specialize an Abstract Data Type - For HL7 messages that include an attribute with an abstract data type, the analyst must select one of the specialized data types appropriate for that abstract type.

d. Add User-Defined Default Values - Source data may not always provide all values required by the HL7 v3 message type. Therefore the analyst defines default values for message elements that won't be mapped due to an absence the source data or that may require a default value if the mapping source data is null.

e. Add a Copy of an Element with Multiple Cardinality - Some clones and attributes may have a maximum cardinality of greater than one. In such cases, the analyst may decide to copy multiple instances of the element in the specification. Later, this will allow the analyst to map different source fields to different copies of the same element in the target.

4.3.7.3 The analyst validates the HL7 v3 specification:

a. Validate the HL7 v3 specification against validation rules.

b. The analyst saves and/or prints the validation results.

4.3.7.4 The analyst saves (or saves as) the HL7 v3 specification to file in an XML format that can be opened by another analyst who can continue working on the specification.

4.3.8 Post Conditions

4.3.8.1 The system successfully manages the HL7 v3 specification without any validation warning or error.

4.3.8.2 The system manages HL7 v3 specification instances with warnings or errors.

4.3.8.3 The system fails to create or open the HL7 v3 specification because of a fatal error.

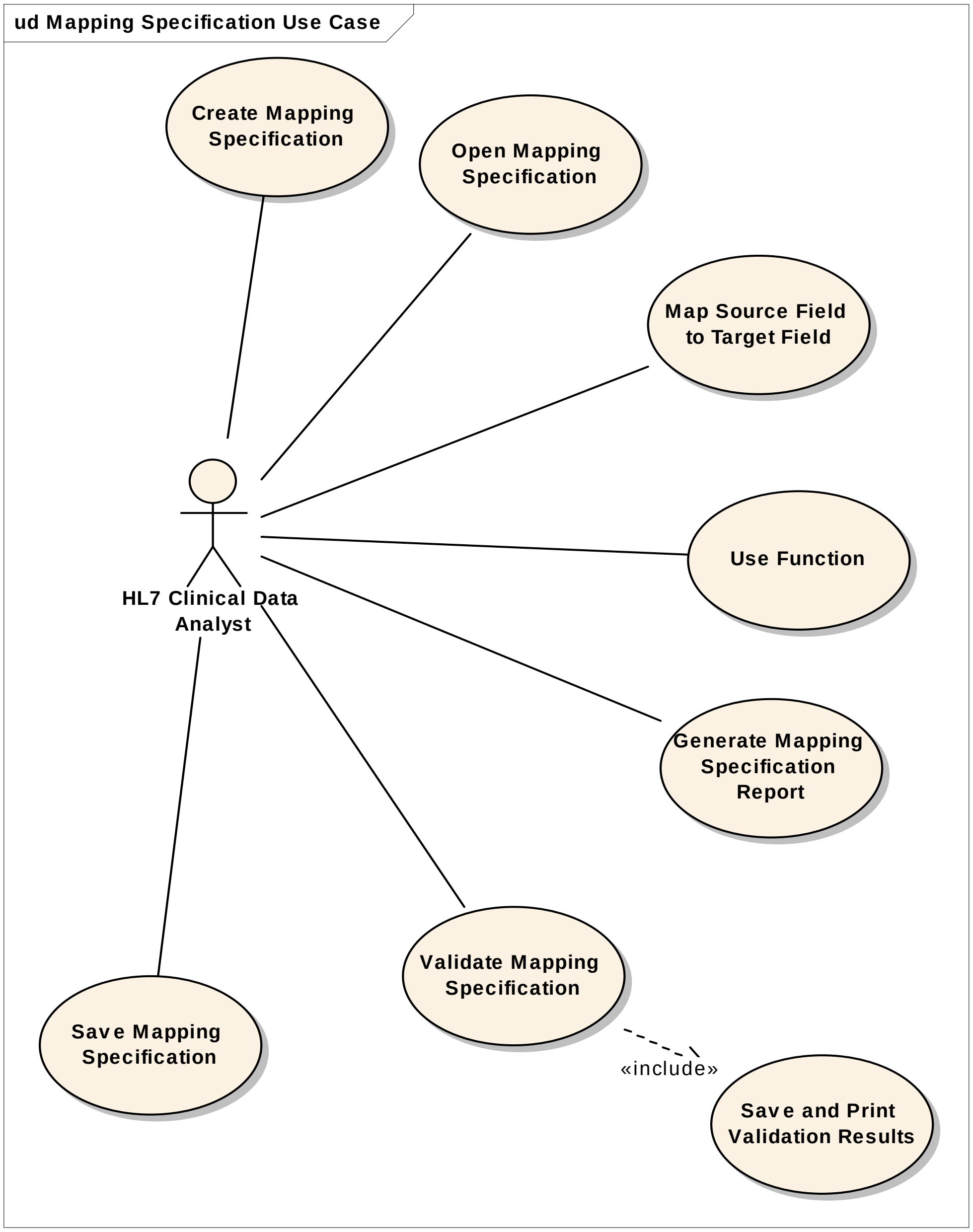
4.3.9 Alternative Courses of Action

4.3.9.1 The system manages the HL7 v3 specification with warnings and/or errors, and then the analyst modifies the specification (returns to step 4.3.7.2).

4.3.9.2 The system fails to create or open the CSV specification. The actor reports the error.

4.4 Mapping Specification Use Case

4.4.1 Use Case Diagram


ud Mapping Specification Use Case
Create Mapping Specification
Open Mapping Specification
Map Source Field to Target Field
Use Function
Generate Mapping Specification Report
Validate Mapping Specification
Save Mapping Specification
«include»
Save and Print Validation Results
HL7 Clinical Data Analyst

4.4.2 Use Case ID: caAdapter_Mapping-UC-04

4.4.3 Brief Description

In this scenario, the HL7 Clinical Data Analyst has an existing CSV specification file and an HL7 v3 specification file that represents the target message type for the data conversion. The Mapping Tool allows the analyst to create a new mapping specification based on the source and target specification, or open an existing mapping specification file which refers to the source and target specification, and then perform “drag-n-drop” operations to map source fields to target fields. The analyst may use functions to manipulate the data on the way from source to target, and save the whole set of mappings to a mapping specification file.

4.4.4 Primary Actor

HL7 Clinical Data Analyst

4.4.5 Preconditions

4.4.5.1 CSV Specification File

The analyst must have a CSV specification file that is an XML-like document describing the format of the CSV file. This includes all segments, their fields in order, and the segment hierarchy.

4.4.5.2 Corresponding HL7 v3 Specification File

To complete this process, the analyst must have a HL7 v3 Specification file corresponding to the type of data in the CSV file from the source system. This file describes the target message type that will be generated from the CSV data.

4.4.6 Basic Flow of Events

4.4.6.1 Create or Open a Mapping Specification File:

a. The HL7 Clinical Data Analyst creates a new mapping file by selecting a CSV specification file as a source and an HL7 v3 specification as a target from the file system.

b. The HL7 Clinical Data Analyst opens an existed mapping file from the file system.

4.4.6.2 The Mapping GUI displays the CSV specification on the source side of the screen and the HL7 v3 specification clones, attributes and fields on the target side of the screen.

4.4.6.3 The analyst may begin the field-to-field mapping by dragging and dropping source fields onto target fields.

4.4.6.4 The analyst may place a particular function on the screen by dragging-n-dropping from the function library. The display of the function includes both input and output parameters as “mappable” items. Then, the user can drag-n-drop source fields onto the input parameters of the function, and drag-n-drop target fields onto the output parameters of the function.

4.4.6.5 The analyst validates the mapping configuration against the mapping validation rules. The analyst may save and/or print the validation message.

4.4.6.6 The analyst may generate reports on the mapping configuration, showing what is mapped from the source perspective or from the target perspective, or a summary of unmapped items from both sides.

4.4.6.7 The analyst saves the mapping specification to the file system in an XML format.

4.4.7 Post Conditions

4.4.7.1 The system successfully manages the Mapping specification without any validation warning or error.

4.4.7.2 The system manages Mapping specification instances with warnings or errors.

4.4.7.3 The system fails to create or open the mapping specification because of a fatal error.

4.4.8 Alternative Courses of Action

4.4.8.1 The system manages the Mapping specification with warnings and/or errors, and the analyst modifies the mapping specification (returns to step 4.4.6.3).

4.4.8.2 The system fails to create or open the mapping specification. The actor reports the error.

5 Data Definitions

The following list of definitions does not include commonly used HL7 terms since the reader is assumed to be familiar with such terms. The list includes terms used by use cases that have been fully documented.

CSV File – A set of lines or rows of comma-separated data values generated from a source clinical system. This file is to be converted into one or more HL7 v3 XML instances using the caAdapter Mapping Tool.

Field – A single data item in a file, separated from other items by commas, except when at the end of a line where a new-line character is the delimiter.

Field Name – A label given to a field, hopefully meaningful in a human-readable way.

Segment Name – The value of the first field in the beginning of a line in a CSV file that serves as an identifier for the type of line.

Segment – A type of a line of data in a CSV file. A segment is comprised of a segment name and one or more fields. Each segment has a unique segment name.

Logical Record – A group of segments that together comprise a meaningful unit of data, such as documentation for a single substance administration event that contains information about the practitioner, the patient, the administered drug, relevant observations, etc. Each CSV file contains one or more logical records.

Root Segment – A type of segment that is always the first segment in a logical record, indicating the start of a new logical record occurrence. Each root segment always occurs only once in a logical record, but other segments may occur zero, one or more times in the logical record.

Segment Hierarchy – The parent-child relationships between the root segment and all other segments in the logical CSV record. Each segment must have a parent segment, except the root segment. There is no limit to how many child segments a parent can have or how many generations a segment hierarchy can have. In a CSV file, each child segment must appear below its parent and above its children (if any). Each child segment may occur zero, one or more times in a logical CSV record indicating a one-to-many relationship between the data in the parent segment and the child segments. Note that segments may be optional and not appear at all in a given logical CSV record.

CSV Specification – The distinct collection of all segments in a given logical record along with their segment names, parent segments, and field names. A CSV specification is displayed in the Mapping Tool GUI using a tree-like hierarchy and represents the structure of data coming from a clinical system that is to be converted into one or more HL7 v3 XML instances.

CSV Specification File – The physical file that contains the XML-like specification describing the CSV file's logical record structure.

Source Validation Results – A set of success/error messages generated from the process of validating a CSV file against a source definition file or the validation rule.

HL7 v3 Specification – A hierarchy of components of an HL7 v3 XML message type, including all elements and attributes down to the data type field level. This hierarchy explicitly includes all levels of recursion required by the nature of the source data.

Mapping – The assignment of a particular source CSV specification field to be converted to a particular target specification field. This is represented in the GUI using a graphical line between elements in the source and target tree hierarchies.

Mapping Specification – The set of all mappings for a given source/target pair. This file is used by the transformation service to effect the actual transformation of an incoming source data to individual XML instances of a given HL7 v3 message type.

Mapping Specification File – The physical file that contains the specification describing a mapping configuration.

6 Sign Off

6.1 Approval

Workspace General Contractor Rep *Print Name* *Date*

Architecture Workspace Rep *Print Name* *Date*

VCDE Workspace Rep *Print Name* *Date*

Workspace Working Group Rep *Print Name* *Date*

NCICB Rep *Print Name* *Date*