



i2b2 User Guide

Import Data View

Table of Contents

INTRODUCTION	3
LAYOUT OF THE IMPORT WIZARD TAB	3
<i>Choose Data Section</i>	4
<i>Verify Data Section</i>	5
Observation Tab	5
Patient Mapping Tab	5
Event Mapping Tab	6
<i>Upload Data Section</i>	7
LAYOUT OF THE IMPORT STATUS TAB	7
<i>Refresh Button</i>	9
LAYOUT OF THE OPTIONS TAB	9
<i>Encryption Fields</i>	10
Encrypt Mapped Numbers	11
Encrypt Text Fields	11
Encrypt Files	11
<i>Create Patient and Visit Number Fields</i>	11
When patient number not found, make a new one	11
When visit number not found, make a new one	11
<i>Number of records per batch</i>	11
ACCESS / SECURITY	12
IMPORT FILES	12
FILE TYPES	12
<i>CSV Files</i>	12
<i>XCEDE Files</i>	13
<i>Database Files</i>	13
DELIMITERS	13
<i>Field Delimiters</i>	13
Text Qualifier	13
<i>End of Line Delimiters</i>	14
DATA TYPES	14
<i>Observation</i>	14
<i>Patient Mapping</i>	16
<i>Event Mapping</i>	16
IMPORT PROCESS	17
IMPORT DATA	17
<i>Choose Data</i>	17
Import Wizard	18
<i>Verify Data</i>	26
<i>Upload Data</i>	27
IMPORT STATUS	28
APPEND FLAG	28
<i>Append Flag = True</i>	28

Introduction

The import data view is a tool for users to upload data into the i2b2 database. The type of data that can be uploaded includes; observations, patients and events.

The import data view is comprised of three different tabs, each of which has a unique function. The **Import Wizard tab** is used to map the data. The **Import Status tab** is used to view the status of the upload batches. The **Options tab** is used to set up the user preferences for the batch being uploaded.

Layout of the Import Wizard Tab

The **Import Wizard tab** is where the data to be imported is mapped, the data is verified and the process of importing is started. The view is comprised of three sections; (1) Choose data, (2) Verify Data and (3) Upload data.

Verify Data Section

The second step is to verify the data being imported is mapped correctly. In this section there are three tabs; (1) Observation, (2) Patient Mapping, and (3) Event Mapping. The type of data being uploaded will determine which tab will be used to verify the data.

OBSERVATION TAB

The **Observation tab** displays the observations to be uploaded. The columns that appear are dependent on the data that is being uploaded.

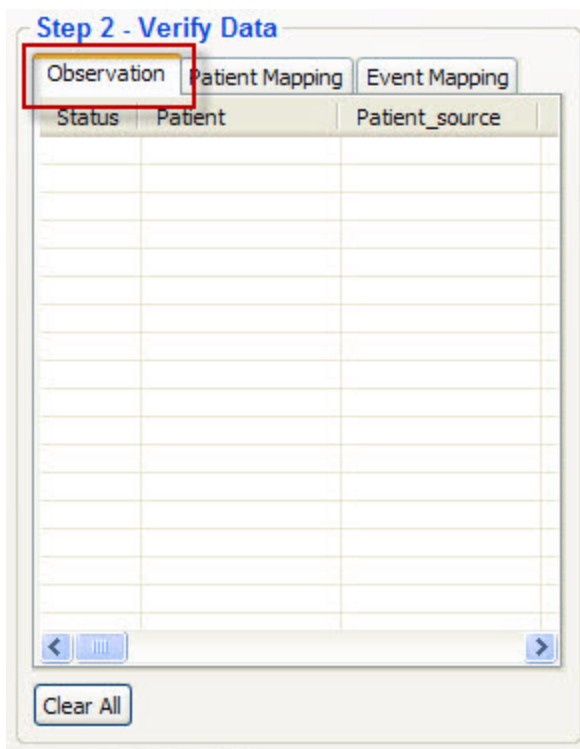


Figure 2 Verify Observations

PATIENT MAPPING TAB

The **Patient Mapping tab** displays the patients to be uploaded. The columns that appear are dependent on the data that is being uploaded.

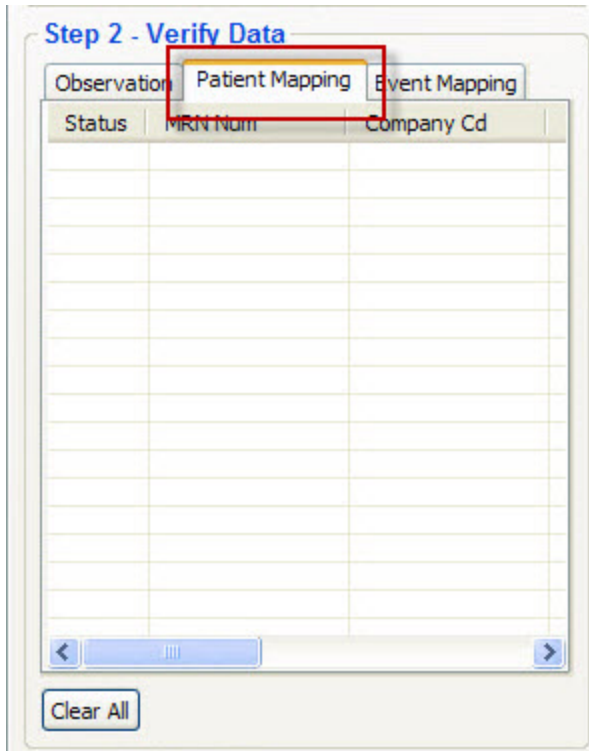


Figure 3 Verify Patients

EVENT MAPPING TAB

The **Event Mapping tab** displays the events to be uploaded. The columns that appear are dependent on the data that is being uploaded.

🔗 *Events are also known as encounters.*

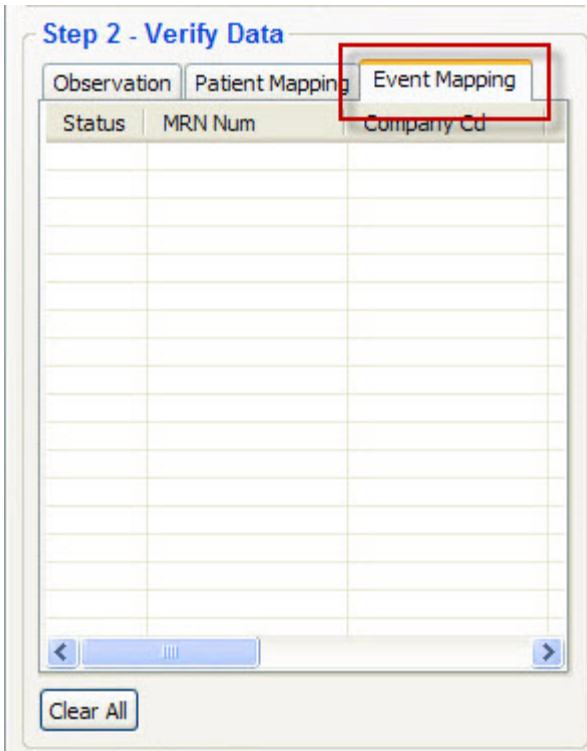


Figure 4 Verify Events

Upload Data Section

The third and final step in the upload process is to start it. This is done by clicking on the upload button in the Upload Data section. A progress bar is displayed and will be activated once the upload process has begun.



Figure 5 Upload Data

Layout of the Import Status Tab

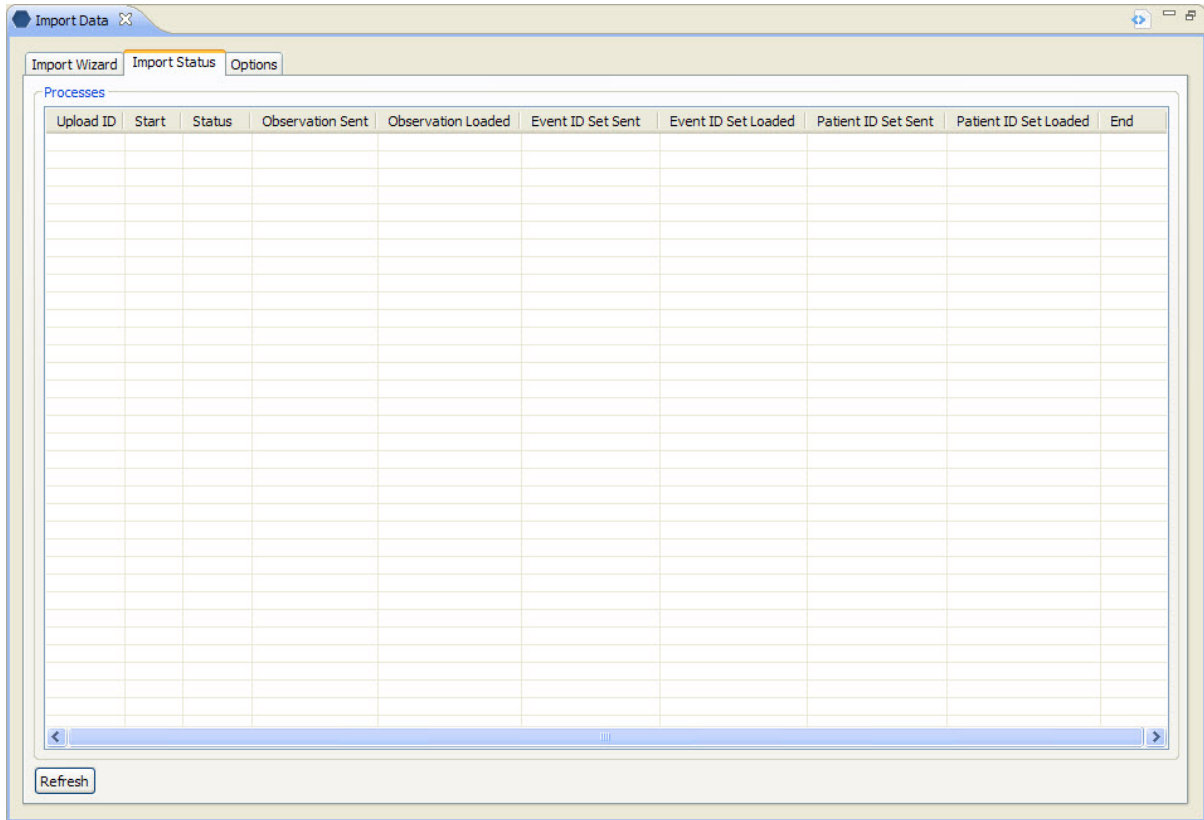


Figure 6 Import Status Tab

Column Name	Description / Use
Upload ID	The unique identifier assigned to the batch when the data was uploaded to the table.
Start	Date the upload process started.
Status	Displays where the batch is in the upload process.
Observation Sent	Displays the number of observation records included in the batch <i>to be</i> uploaded.
Observation Loaded	Displays the number of observation records in the batch that <i>were</i> loaded into the table.
Event ID Set Sent	Displays the number of event records included in the batch <i>to be</i> uploaded.
Event ID Set Loaded	Displays the number of event records in the batch that <i>were</i> loaded into the table.
Patient ID Set Sent	Displays the number of patient records included in the batch <i>to be</i> uploaded.

Patient ID Set Loaded	Displays the number of patient records in the batch that <i>were</i> loaded into the table.
End	Date the upload process ended.

Refresh Button

Clicking on the Refresh button will force the view to refresh and display the latest status of the upload batches.

Layout of the Options Tab

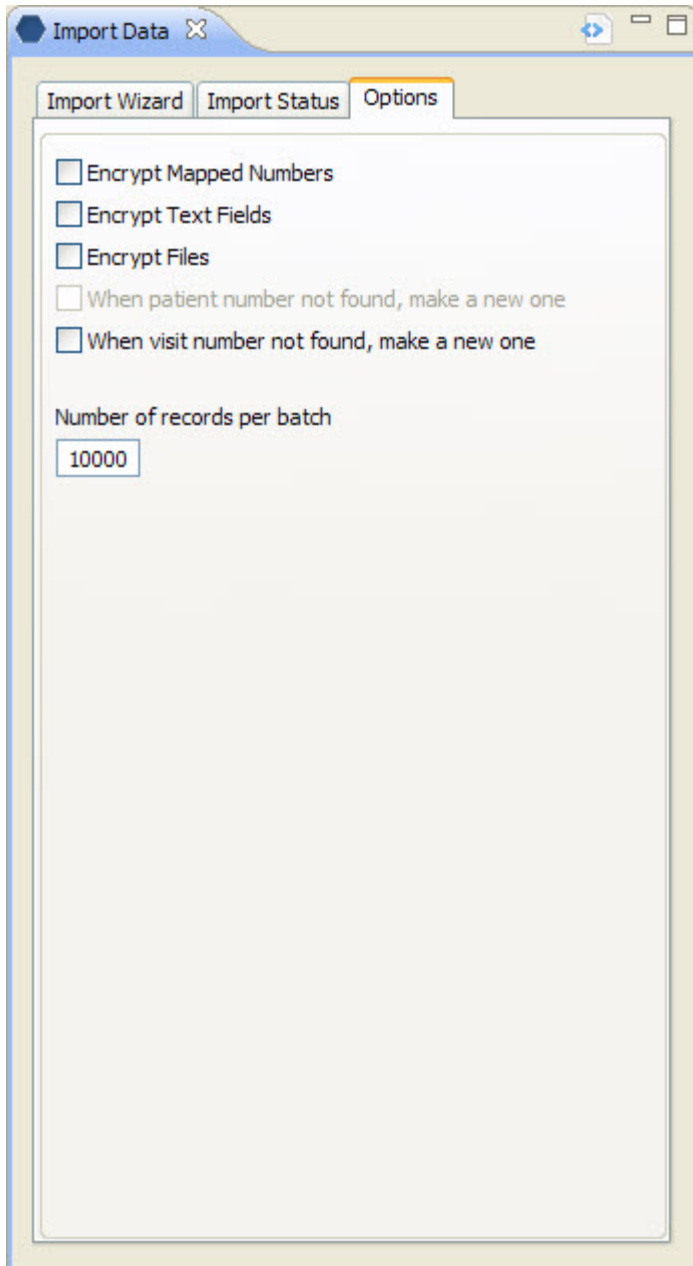


Figure 7 Options Tab

Encryption Fields

In order to comply with confidentiality standards, patient identifying data must be encrypted. There are three areas that can be encrypted; (1) Mapped Numbers, (2) Text Fields and (3) Files.

ENCRYPT MAPPED NUMBERS

Mapped numbers are the number fields that *link* information in one table to another table. Some examples are the **patient_num** field and the **encounter_num** field.

ENCRYPT TEXT FIELDS

Text Fields that are encrypted are the blob data type. An example of what may be in a text field is a patient note or report for a radiology exam.

ENCRYPT FILES

Create Patient and Visit Number Fields

When the data being loaded does not have an existing patient and/or encounter number, users can choose to have the system randomly generate one.

WHEN PATIENT NUMBER NOT FOUND, MAKE A NEW ONE

A check mark at this field means the system will randomly generate an alphanumeric ID for the patient_id and assign the patient_source for the new ID.

WHEN VISIT NUMBER NOT FOUND, MAKE A NEW ONE

A check mark at this field means the system will randomly generate an alphanumeric ID for the encounter_id and assign the encounter_source for the new ID.

Number of records per batch

This field is used to specify the number of records to be uploaded in a single batch. If the number of records in a file exceeds the number of records defined per batch, then the system will continue to generate additional batches until all the records are uploaded.

Example:

Number of records per batch: 10000

Number of records in the import file: 35000

In this example, the system will create a total of 4 batches. The first three batches will contain 10000 records and the fourth one will contain the remaining 5000 records.

Access / Security

The import feature is only available to those users who have been set up with the role of manager.

 **Roles are set up in the i2b2 administration module.**

Import Files

This section contains information about the import files. It describes the different types of files, delimiters and the data types.

File Types

In the import tool there are three different file types listed. At this time only one is available for use. Although only one is in use at this time, all three are defined below.

CSV Files

Comma-separated values (CSV) file is used when importing text files. Each line in the file corresponds to a row in a table. The fields in each line are separated by a comma or some other delimiter. Each field represents a column in the table.

Example: Simple CSV file containing patient information.

File format:

```
patient_id,patient_id_source,patient_id_status,patient_num,sourcesystem_cd
123456,BWH,A,000123,BICS
987654,MGH,A,000004,PACS
```

Table format:

patient_id	patient_id_source	patient_id_status	patient_num	sourcesystem_cd
123456	BWH	A	000123	BICS
987654	MGH	A	000004	PACS

XCEDE Files

XML-Based Clinical Experiment Data Exchange Schema (XCEDE) files are an extensive metadata hierarchy for describing and documenting research and clinical studies.

 ***At this time, the XCEDE file format is not available for uploading.***

Database Files

Database files are a particular file format that is specific to the databases in use.

 ***At this time, the Database file format is not available for uploading.***

Delimiters

In a CSV file, the **delimiter** is a sequence of one or more characters. These characters are used to identify the boundaries between different sections of a text file.

Field Delimiters

The **field delimiters** are specific to a single line in a file. This type of delimiter is used to designate the columns for each row. The following field delimiters are available in the i2b2 Import Data view.

1. Tab ()
2. Semicolon (;)
3. Comma (,)
4. Space ()
5. Pipe (|)

TEXT QUALIFIER

The **Text Qualifier** is a character used to indicate that everything in between it should be imported exactly as it appears. In the i2b2 Import Data View there are two options currently available.

1. Double Quote (“)
2. Single Quote (‘)

ⓘ *At the Text Qualifier field in the import wizard the drop down list will also contain {none}, which means the file does not contain any text qualifiers.*

End of Line Delimiters

The **end-of-line delimiters** are used to designate the end of a line of text. These delimiters are also known as a **line break**.

Data Types

Observation

Observations are collections of phenotypic data and may contain values associated with a concept, such as a value of the systolic blood pressure. Other observations may include lab results, notes, and reports.

ⓘ *These fields are based on the i2b2 Harvard Demo database. Fields and their usage may differ for individual sites.*

Column Name	Description / Use
concept_cd	Unique id for the concept.
confidence_num	Currently not used.
destination_file	Currently not used.
download_date	Date the data was downloaded from the source system.
encounter_num	Unique id assigned by the source system.
encounter_source	The source system that assigned the encounter_num.
end_date	Date the observation ended.


import_date	Date the record was imported.  <i>The date is automatically assigned when the record is imported.</i>
location_cd	Place the observation was made.
modifier_cd	Modifier assigned in the source system
nval_num	Numerical value (i.e. lab test result).
observation_blob	Blob of text (i.e. report)
original_file	Name of the file the data was imported from. Can be used for reference purposes.
patient_num	Unique number assigned to the patient.
patient_source	The source system that assigned the patient_num.
provider_id	Unique id for the provider associated to the observation.
quantity_num	Numerical value representing the quantity.
sourcystem_cd	Unique id for the system that is the source for the record.
start_date	Date the observation began.
tval_char	Text value (i.e. Lab/Micro result)
units_cd	Units associated to the value (i.e. mil/cmm)
update_date	Date the record was last updated.
valtype_cd	The type of value.
valueflag_cd	Value flag (i.e. high, low, abnormal)

Required Fields: Observation

The following fields are required when importing an observation into the i2b2 database.

- concept_cd
- start_date
- encounter_source
- patient_source

Patient Mapping

Column Name	Description / Use
download_date	Date the data was downloaded from the source system.
import_date	Date the record was imported.  <i>The date is automatically assigned when the record is imported.</i>
patient_id	Unique id for the patient. This id is assigned by the source system.
patient_id_source	The source system that assigned the patient_id.
patient_id_status	The status of the patient_id (i.e. active, inactive, merged)
patient_num	Unique number assigned to the patient.
sourcesystem_cd	Unique id for the system that is the source for the record.
update_date	Date the record was last updated.


Required Fields: Patient

The following fields are required when importing a patient record into the i2b2 database.

- patient_id
- patient_id_source

Event Mapping

Column Name	Description / Use
download_date	Date the data was downloaded from the source system.
encounter_id	Unique id for an encounter. This id is assigned by the source system.
encounter_id_source	The source system that assigned the encounter_id.
encounter_id_status	The status of the encounter_id (i.e. active)
encounter_num	Unique id assigned by the source system.

import_date	Date the record was imported.  <i>The date is automatically assigned when the record is imported.</i>
patient_id	Unique id for the patient. This id is assigned by the source system.
patient_id_source	The source system that assigned the patient_id.
sourcesystem_cd	Unique id for the system that is the source for the record.
update_date	Date the record was last updated.

Required Fields: Event (encounter)

The following fields are required when importing an encounter into the i2b2 database.

- encounter_id
- encounter_id_source
- patient_id
- patient_id_source

Import Process

This section contains information about the process of importing files. It provides step by step documentation on using the import wizard, starting the upload and how to check the status of the file being uploaded. Documentation on the functionality of the Append Flag is also contained in this section.

Import Data

There are basically three steps to importing data into the i2b2 database. (1) Choose the data, (2) verify the data and (3) start the import.

The Import Data View provides a wizard to assist users in choosing and mapping the data to the correct fields in the i2b2 database.

Choose Data

IMPORT WIZARD

The import wizard is designed to take you through each step of the import process.

1. Click on the **Start File Import Wizard** button located in the Choose Data section.

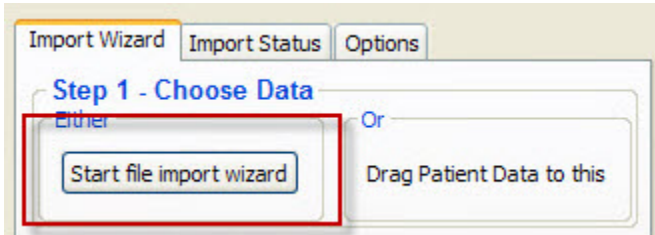


Figure 8 Start File Import Wizard

2. The import wizard window will open to the first page.

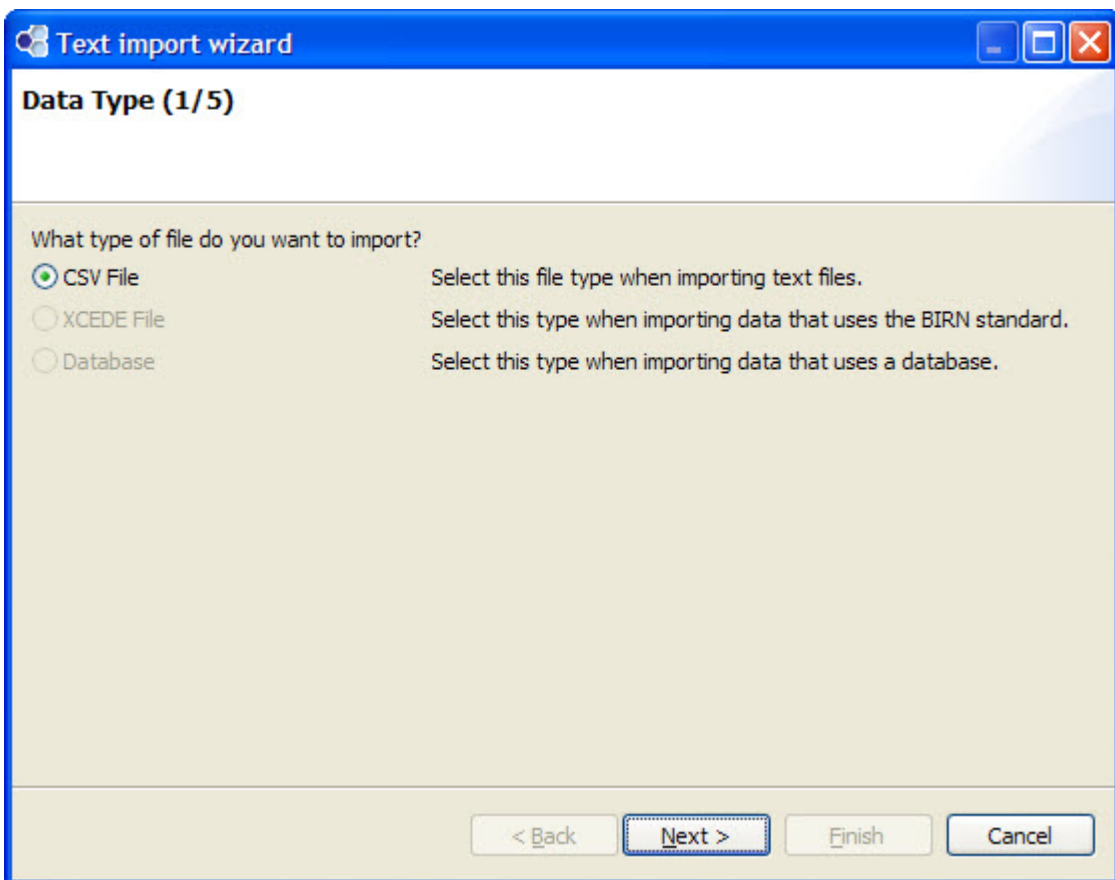


Figure 9 Import Wizard Page 1

3. Select the type of file you are importing.
4. Click on the **Next** button located at the bottom of the window.

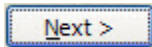


Figure 10 Next Button

5. A dialog box will open.

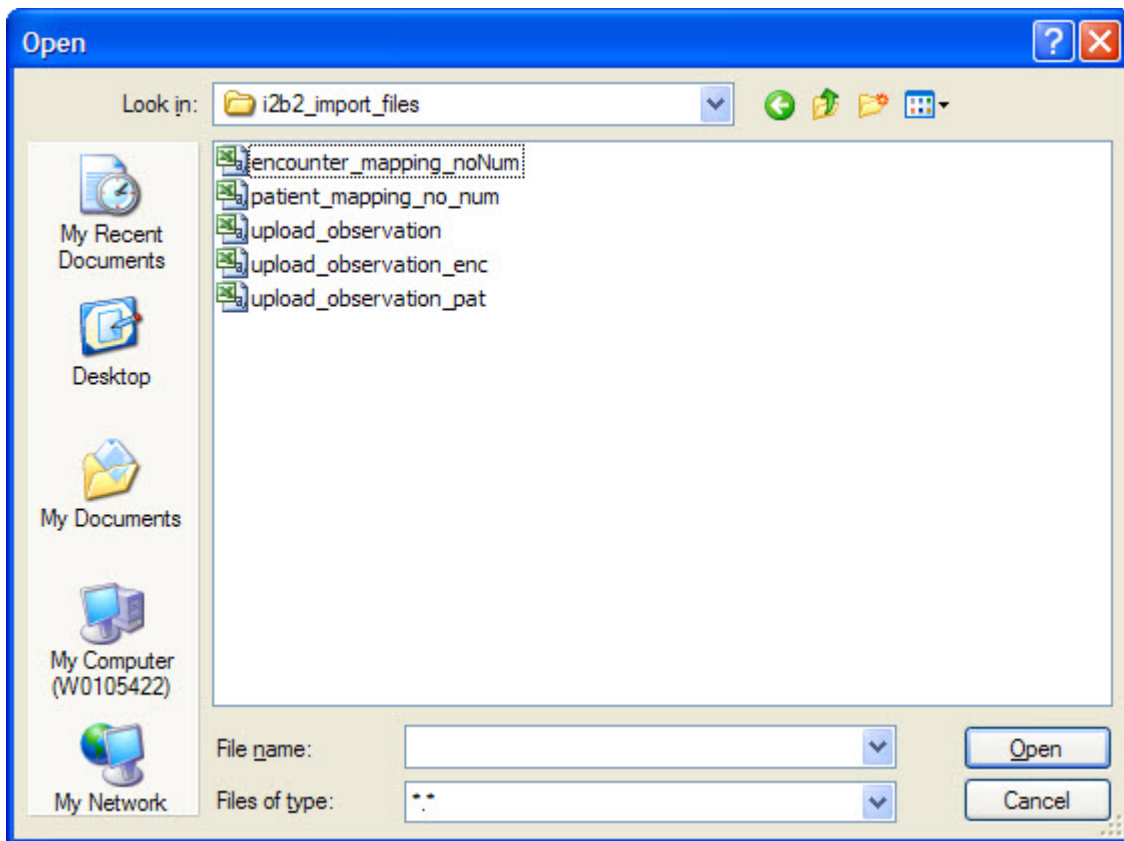


Figure 11 Select File Dialog Box

6. Select the file to be imported.
7. Click on the **Open** button.

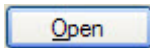


Figure 12 Open Button

8. The second page of the wizard will now open.

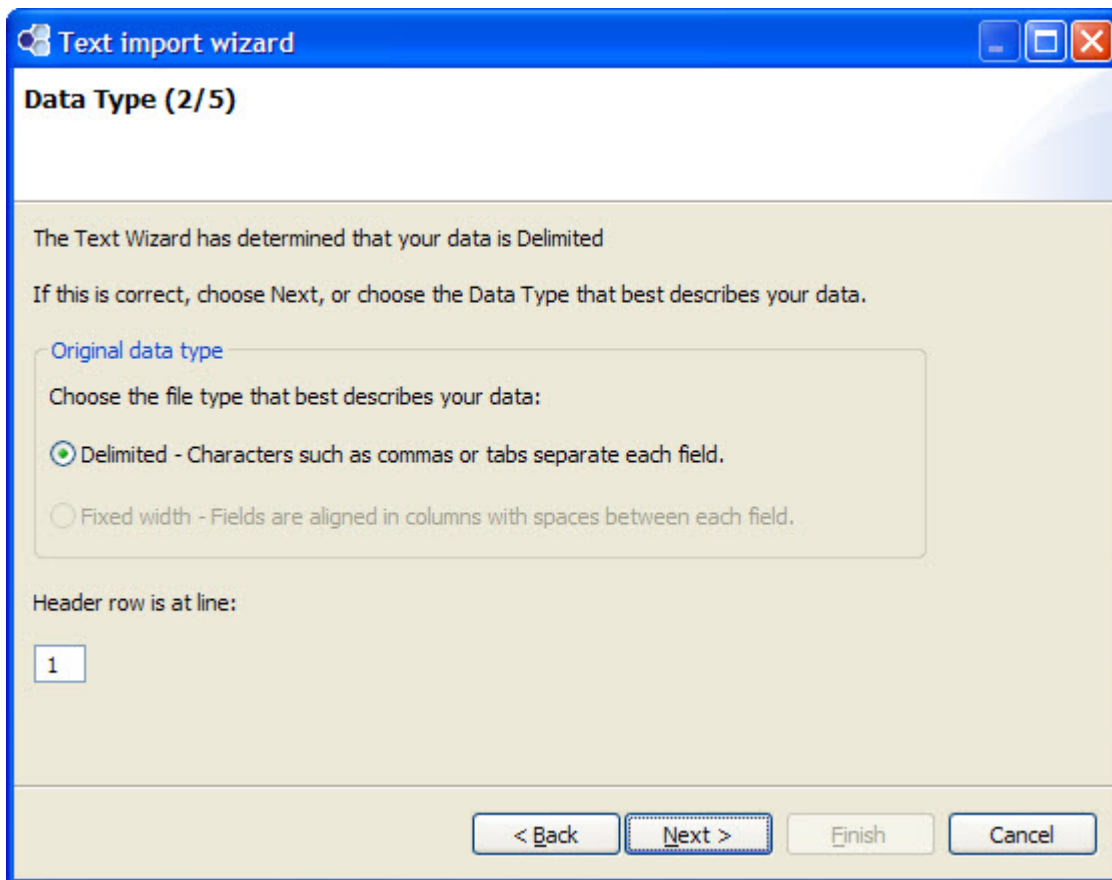


Figure 13 Import Wizard Page 2

9. Select the type of file being imported.

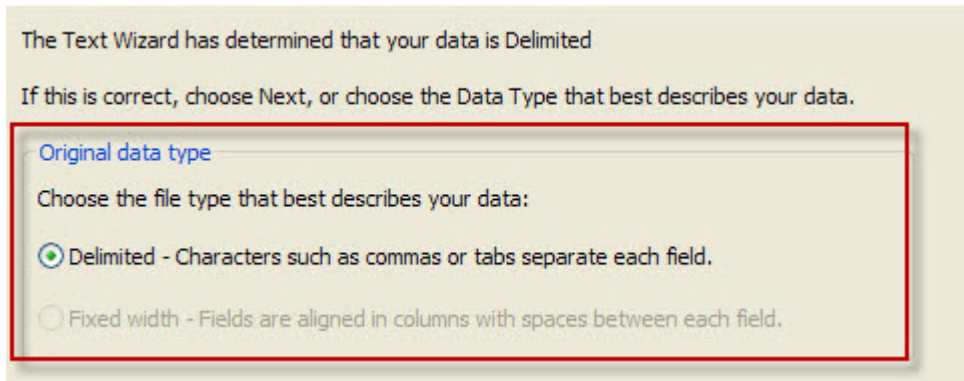


Figure 14 File Type Selection

10. Enter the correct line number at the Header row is at line prompt.

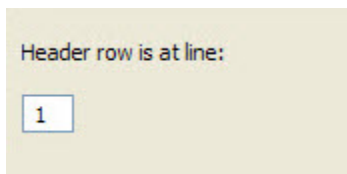


Figure 15 Header Row Selection

11. Click on the **Next** button located at the bottom of the window.

12. The third page of the wizard will be displayed.

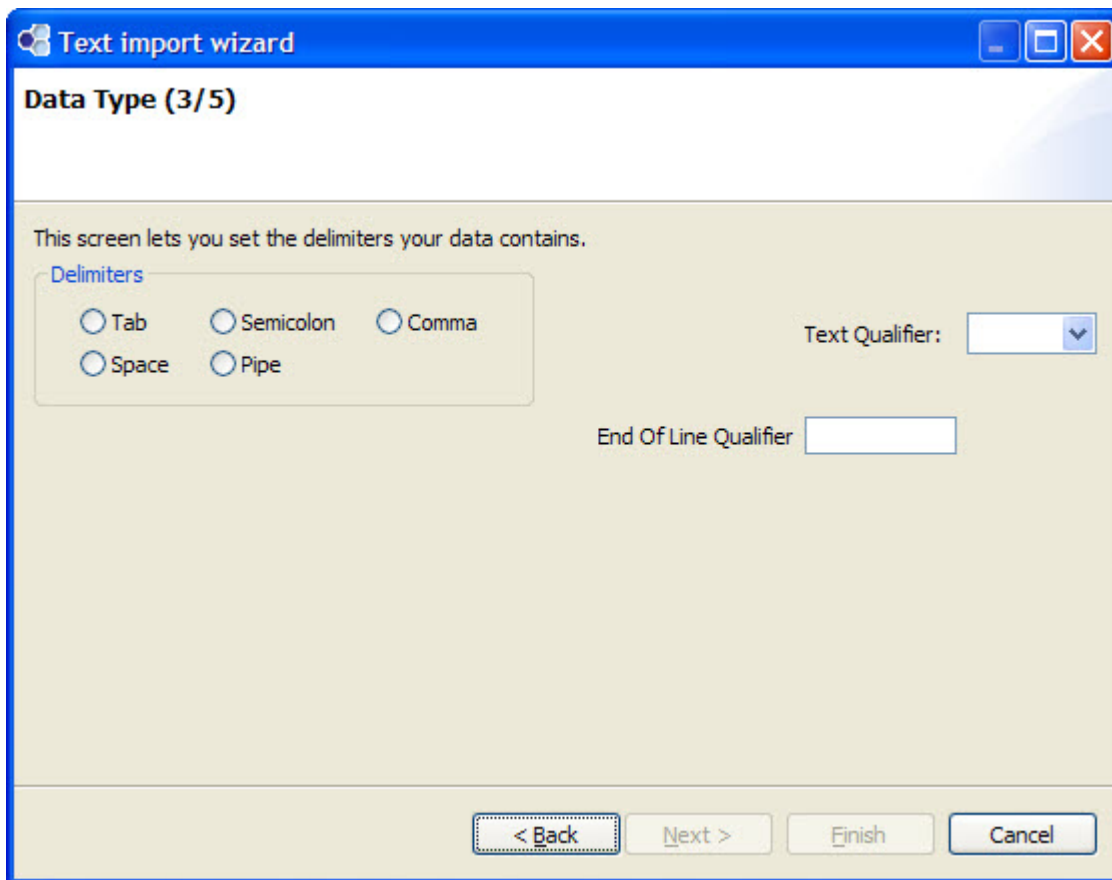


Figure 16 Import Wizard Page 3

13. Check off the appropriate delimiter.
14. If appropriate, enter a qualifier at the Text Qualifier field.
15. If appropriate, enter a qualifier at the End of Line Qualifier.
16. Click on the **Next** button.
17. The fourth page of the wizard will be displayed.

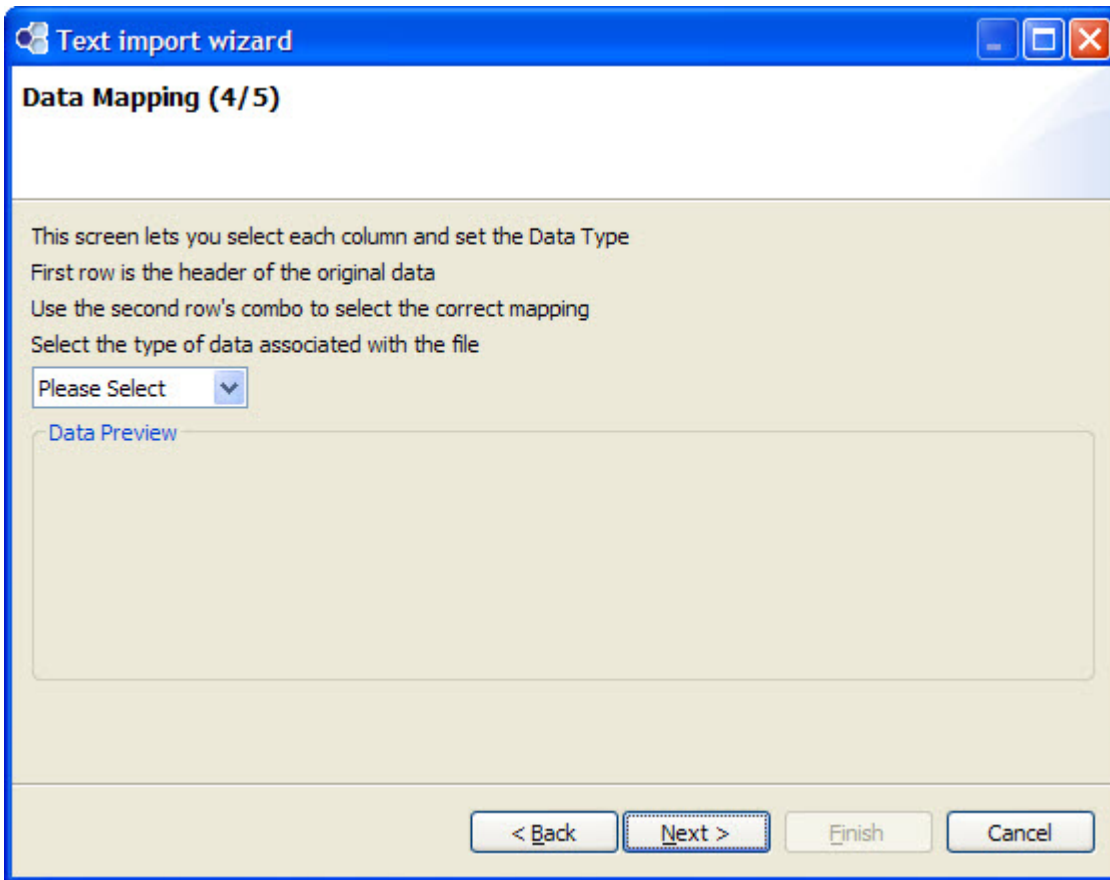


Figure 17 Import Wizard Page 4

18. Select the type of data associated with the file.

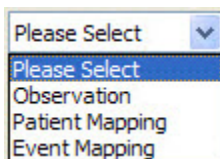


Figure 18 Type of Data Selection

19. The data in the file will appear in the Data Preview Section

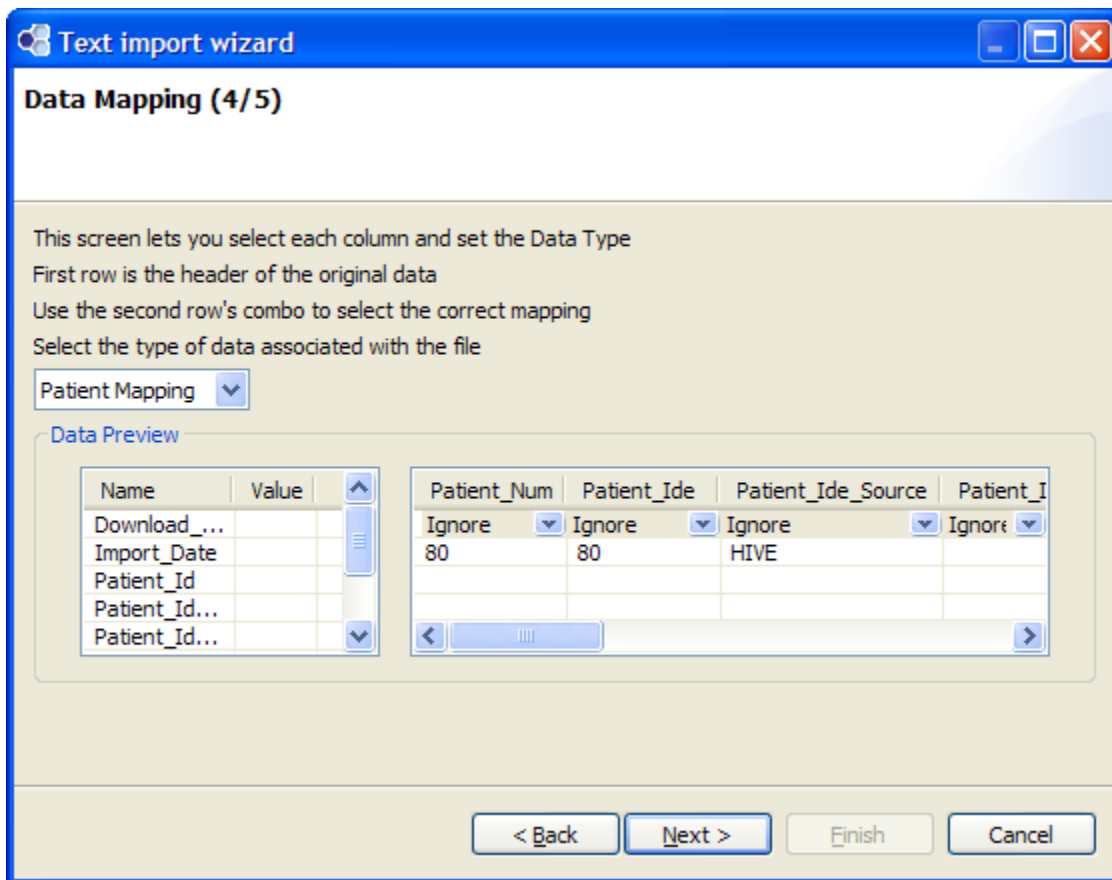


Figure 19 Data Preview on Data Mapping Page

20. Verify the data in the columns matches the column heading (row 1).
21. If the data does not match click on the column heading in row 2.
22. A drop down list will display.

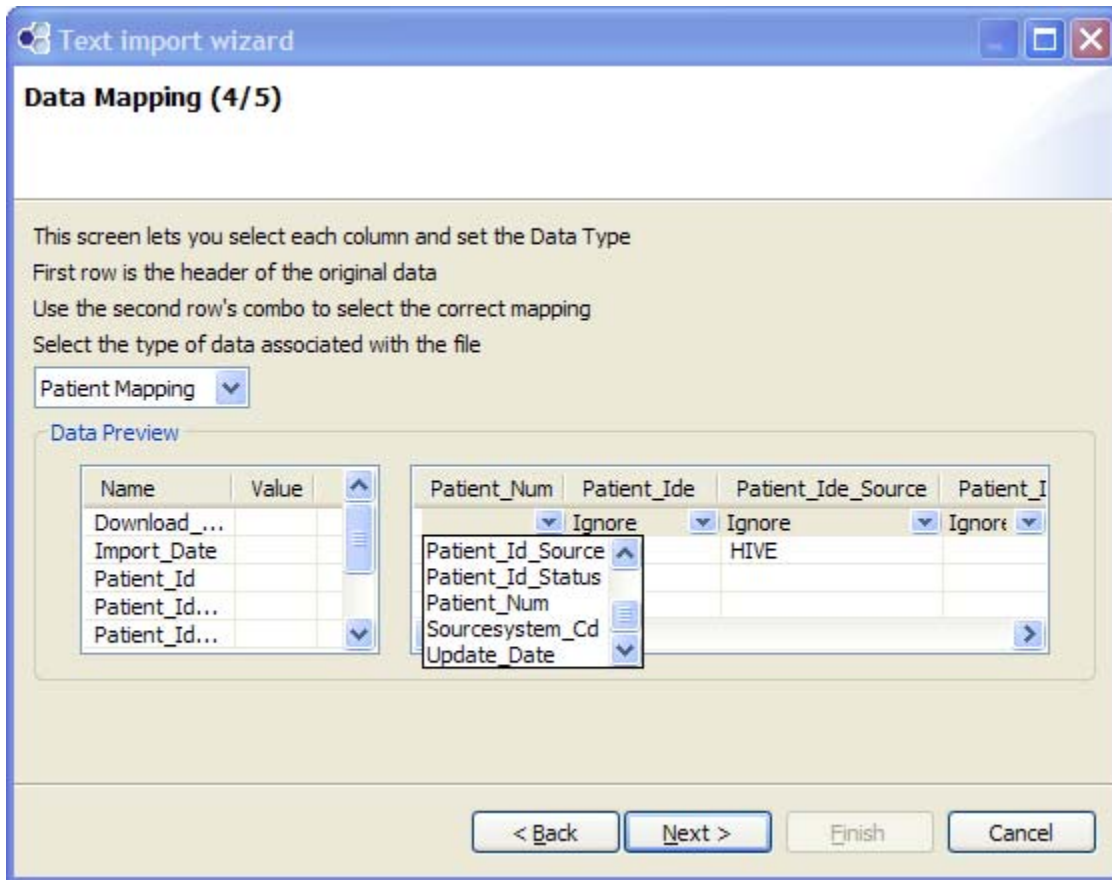


Figure 20 Mapping Selections

23. Select the appropriate column heading.
24. The left side of the Data Preview section is where you can set a default value for a particular column. For instance, if the Source System for all your records is the same you would enter that value on the left hand side at the SourceSystem_CD field.

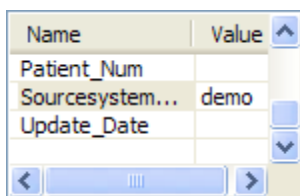


Figure 21 Default Value Setting

25. Click on the **Next** button.
26. The fifth page of the wizard will be displayed.

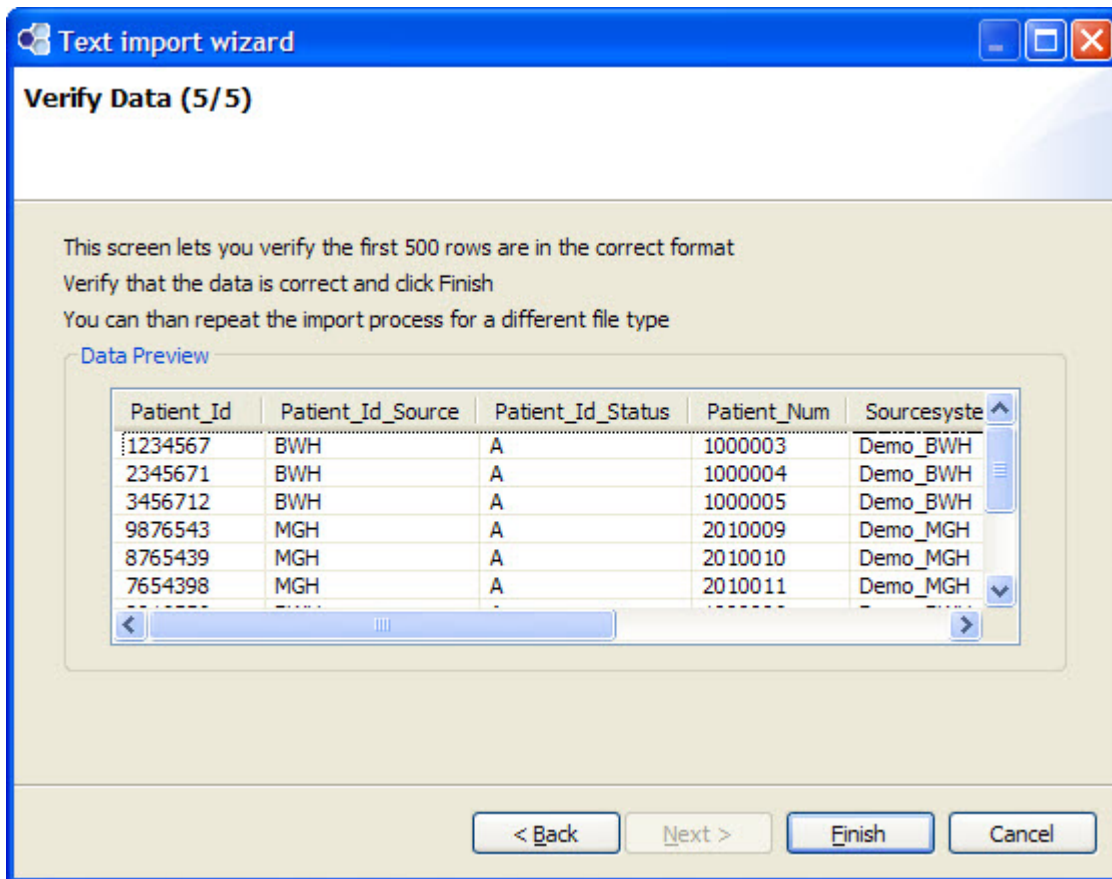


Figure 22 Import Wizard Page 5

27. Scroll through the records to verify the data for each row is displayed in the appropriate columns.
 - a. If the data is incorrect, click on the **Cancel button** and update the file with the appropriate delimiters.
 - b. If the data is correct, click on the **Finish button**.
28. Once you have finished mapping the data with the import wizard the system will return you to the **Import Data view** where all the records in the file will appear in the Verify Data section.

Verify Data

The second step in the import process is to verify the data being imported is correct. Once you have finished the wizard, the data in your file will now appear under the appropriate tab in the Verify Data Section. Scroll through the data to verify everything appears correctly.

Import Status

Once the upload process has begun, a user can go to the Import Status tab to view the status of the batch. This view will display the batch number, how many records were in the file and how many were imported.

Append Flag

Assumption: the record(s) in the update file (new record) has the same primary key as a record(s) in the associate table (existing record).

Primary Key includes:

- Encounter number
- Patient number
- Concept code
- Start date
- Modifier code
- Observer code

Append Flag = True

Following conditions will result in the new record **replacing** the existing record:

new record update date	equal to (=)		update date on the existing record	
new record update date	greater than (>)		update date on the existing record	
new record update date	is not null	AND	update date on the existing record	null
new record update date	null	AND	update date on the existing record	null

Following conditions will result in **ignoring** new record and **not** updating the existing record:

new record update date	less than (>)		update date on the existing record	
new record update date	null	AND	update date on the existing record	is not null

