

ExportXLS v3.3 User's Manual

Introduction

This i2b2 Webclient plugin enables the user to obtain the specified concepts for a set of patient. The user can then further capture those data in a XLS (Excel spreadsheet Workbook) or CSV (comma-separated-values) file.

This plugin has been quite popular and has undergone several revisions, by several institutions, each contributing to its robustness, usability and user-friendliness.

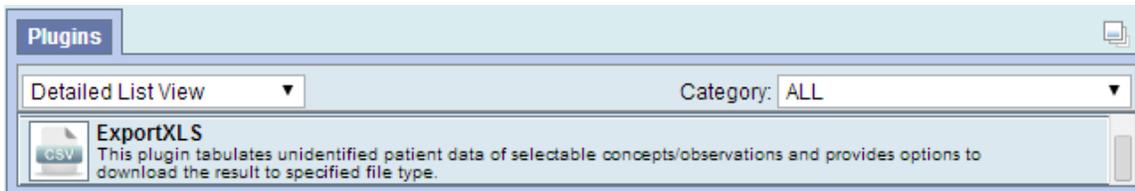
The following are some of the key functions in this version of this plugin:

- Tabulates any selected concepts (ontology terms) relevant to (observed on) a set of selected patients.
- Optionally uses smaller queries of patient subgroups (from the large patient set) iteratively, in place of one single large query of that huge patient set, to reduce risk of overwhelming the server (i2b2 hive).
- Optionally uses selected subset of patients from a large patient set, to further reduce risk of overwhelming the server.
- Provides feedbacks on current progress, elapsed time, and estimated remaining run time.
- Provides notification & suggestion in cases of paging (query-subgrouping) by the server.
- Provides termination, indication & suggestion (for recovery) when the server failed.
- Provides detailed 'HELP' and 'HINT' dialogs in response to corresponding hotlinks strategically placed next to input fields that may need explanations or suggestion.
- Offers different formats and levels of detail.
 - Optionally returns appropriate 'Observation BLOBs' when 'valuetype_cd' is 'B' and the user has the appropriate access right ("DATA_DEID").
- Optionally replaces patient IDs with ascending numbers.
- Minor enhancements & bug fixes in certain demographic patient details.

This document outlines the usage of this plugin. Please refer to its complimentary document, the "ExportXLS v3.3 Installation Guide", for installation details.

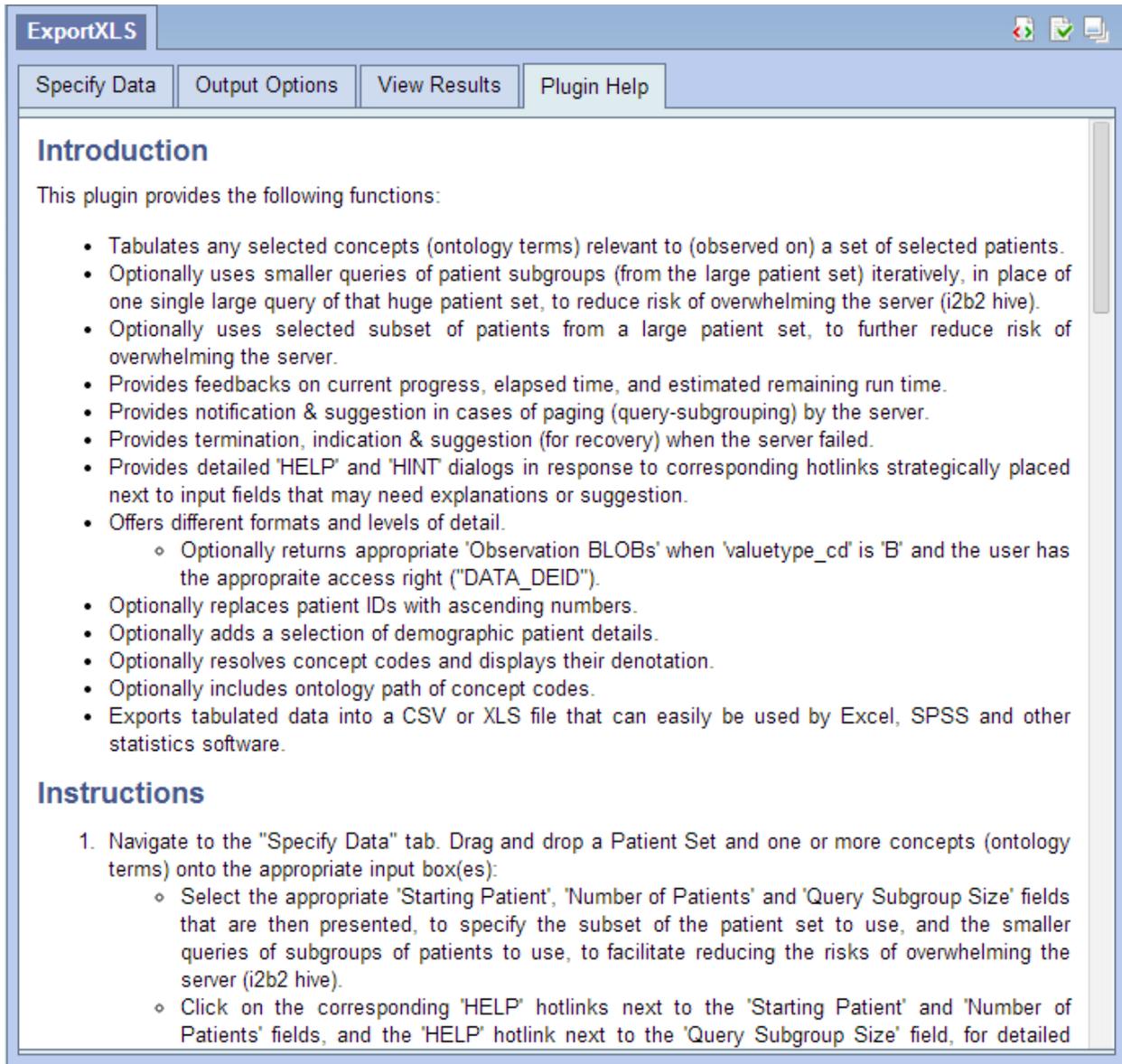
Instructions

As with any i2b2 Webclient plugins, this plugin is listed under the "Plugins" panel under the "Analysis Tools" tab.



The “Plugin Help” Tab

The first thing a user should do, if this document is not readily available, is to click its “Plugin Help” tab, which contains a summary of the contents in this document.



The screenshot shows a software window titled "ExportXLS" with a menu bar containing "Specify Data", "Output Options", "View Results", and "Plugin Help". The "Plugin Help" tab is active, displaying the following content:

Introduction

This plugin provides the following functions:

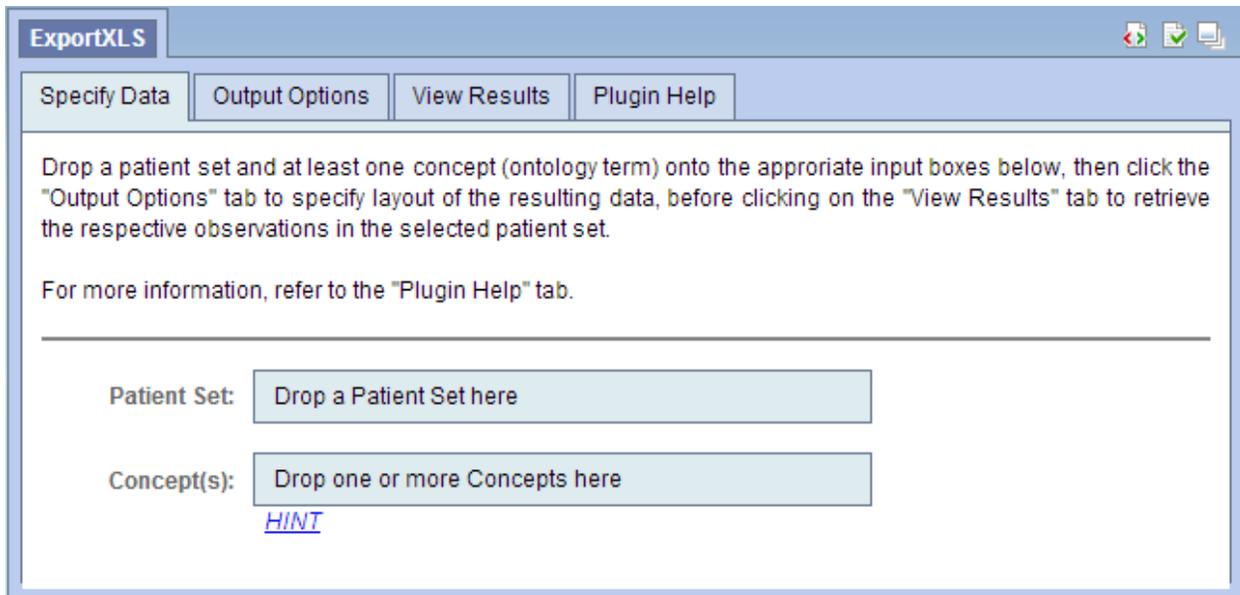
- Tabulates any selected concepts (ontology terms) relevant to (observed on) a set of selected patients.
- Optionally uses smaller queries of patient subgroups (from the large patient set) iteratively, in place of one single large query of that huge patient set, to reduce risk of overwhelming the server (i2b2 hive).
- Optionally uses selected subset of patients from a large patient set, to further reduce risk of overwhelming the server.
- Provides feedbacks on current progress, elapsed time, and estimated remaining run time.
- Provides notification & suggestion in cases of paging (query-subgrouping) by the server.
- Provides termination, indication & suggestion (for recovery) when the server failed.
- Provides detailed 'HELP' and 'HINT' dialogs in response to corresponding hotlinks strategically placed next to input fields that may need explanations or suggestion.
- Offers different formats and levels of detail.
 - Optionally returns appropriate 'Observation BLOBs' when 'valuetype_cd' is 'B' and the user has the appropriate access right ("DATA_DEID").
- Optionally replaces patient IDs with ascending numbers.
- Optionally adds a selection of demographic patient details.
- Optionally resolves concept codes and displays their denotation.
- Optionally includes ontology path of concept codes.
- Exports tabulated data into a CSV or XLS file that can easily be used by Excel, SPSS and other statistics software.

Instructions

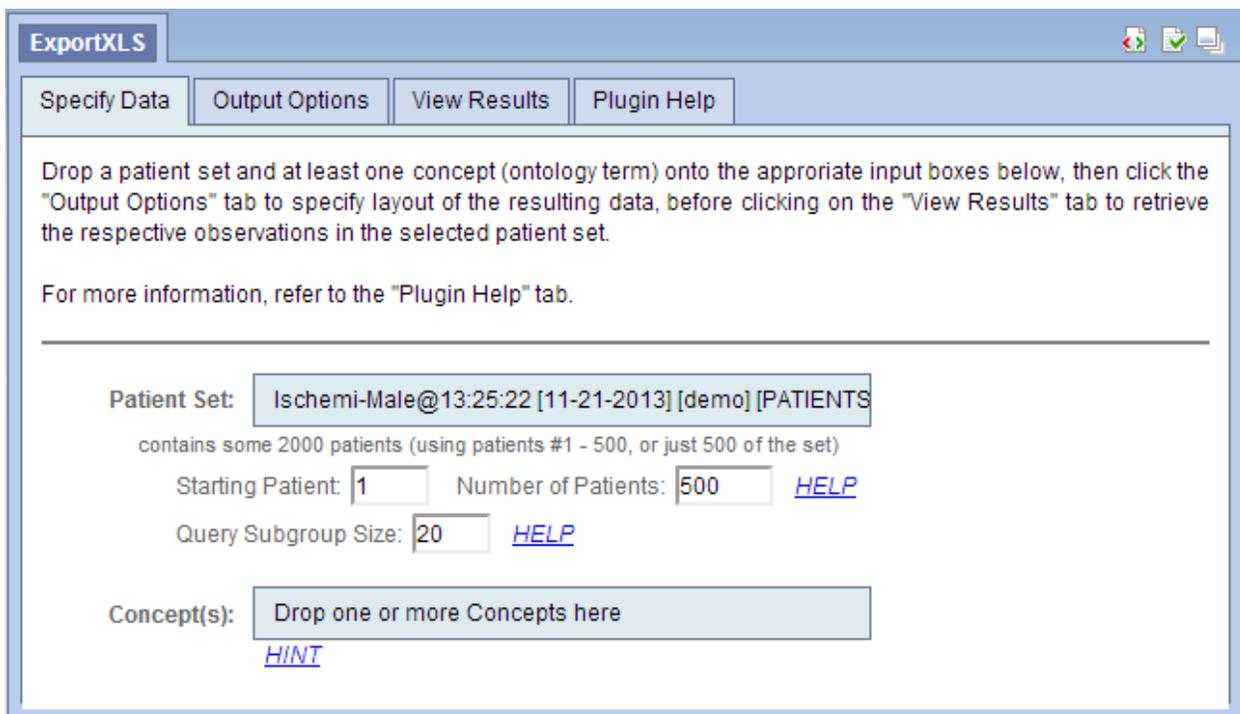
1. Navigate to the "Specify Data" tab. Drag and drop a Patient Set and one or more concepts (ontology terms) onto the appropriate input box(es):
 - Select the appropriate 'Starting Patient', 'Number of Patients' and 'Query Subgroup Size' fields that are then presented, to specify the subset of the patient set to use, and the smaller queries of subgroups of patients to use, to facilitate reducing the risks of overwhelming the server (i2b2 hive).
 - Click on the corresponding 'HELP' hotlinks next to the 'Starting Patient' and 'Number of Patients' fields, and the 'HELP' hotlink next to the 'Query Subgroup Size' field, for detailed

The “Specify Data” Tab

Navigate to the "Specify Data" tab. Drag and drop a Patient Set and one or more concepts (ontology terms) onto the appropriate input box(es):



As soon as a “Patient Set” is specified (i.e. dropped in), then more fields will be displayed for additional input:



Select the appropriate 'Starting Patient' (defaulted to 1), 'Number of Patients' (defaulted to 500) and 'Query Subgroup Size' (defaulted to 20) fields that are then presented, to specify the subset of the patient set to use, and the smaller queries of subgroups of patients to use, to facilitate reducing the risks of overwhelming the server (i2b2 hive).

Click on the corresponding 'HELP' hotlink next to the 'Starting Patient' and 'Number of Patients' fields, and the 'HELP' hotlink next to the 'Query Subgroup Size' field, for detailed information (in corresponding pop-up dialogs); and enter corresponding appropriate values.

About 'Starting Patient' and 'Number of Patients':

If your patient set contains thousands of patients, you may want to use just a subset of that. By so doing, you will speed up your query as well as reduce the likelihood of overwhelming the server (i2b2 hive) to the point of failure (and no result).

Also, if you already encountered a server failure (overwhelmed by the combination of large number of patients times lots of concepts), then you may like to rerun this plugin several times. each time specifying different 'Starting Patient' and 'Number of Patients'.

For instance, use 1 and 500, respectively, in your first run; follow with 501 and 500, respectively, in your second run; then 1001 and 500, respectively, in your third run; and so on, until you get enough data.

The example above should also show that the 'Starting Patient' and 'Number of Patients' refer to the patient entry order in the patient set only, which has no bearing on the actual patient IDs.

In addition, if the sum of the 'Starting Patient' and the 'Number of Patients' values exceeds the total patient count, then it'll be adjusted to the remaining patient count accordingly.

OK

About 'Query Subgroup Size':

Please note that a large query (i.e. a large patient set with many concepts, especially ones consisting many subfolder trees) may overwhelm the server (i2b2 hive), which may fail (where no data can be rendered at all) after considerable delays and timeouts. To avoid this problem, you may want to set the 'Query Subgroup Size' value.

Setting the 'Query Subgroup Size' instructs this plugin to temporarily divide up your patient set into subgroups of your specified size, and then iteratively make request of relevant data for each of these resulting subgroups, and then collate and render the returned data (when all these smaller queries are completed).

Since each subgroup should be of small enough size, the requests hopefully would not cause the server to fail or hang.

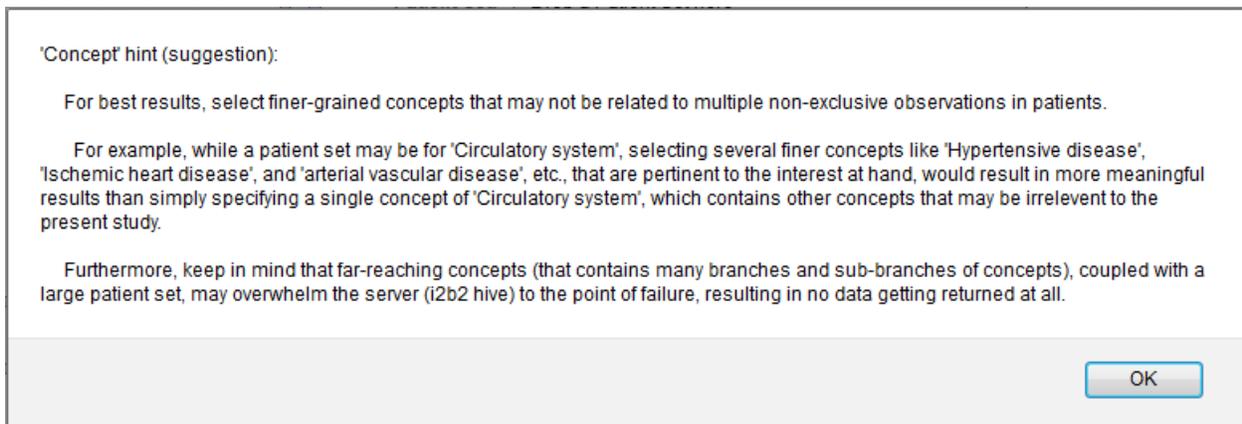
The ideal 'Query Subgroup Size' value cannot be predicted in general, and strongly depends on the number of observations (related to the total numbers of concepts and their complexities) returned; but values of 20 to 50 may be good. A higher value may result in faster processing but may also carry higher risk of the server failing (where no data can be rendered at all).

Incidentally, setting a value of 0 instructs this plugin not to divide up the original query into smaller queries (and carries the risk of overwhelming and failing the server, as well as practically no 'Elapsed time' and 'estimated remaining run time' updates).

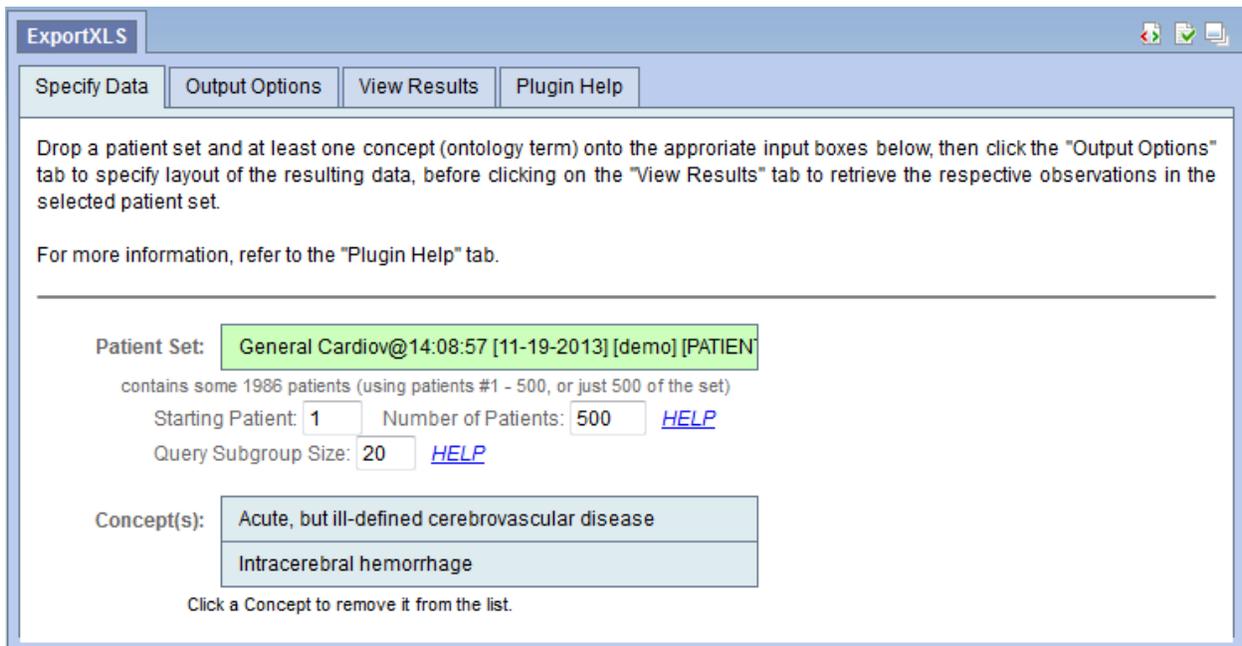
OK

Note that the sample pop-up dialogs are from a Firefox browser, different browsers may have different style dialogs (that are similar but contain the same key contents).

The following pop-up dialog will be displayed when the “*HINT*” hotlink, below the “Concept(s)” input box, is clicked, to provide suggestions regarding concept specifications that may result in faster turn-around time:

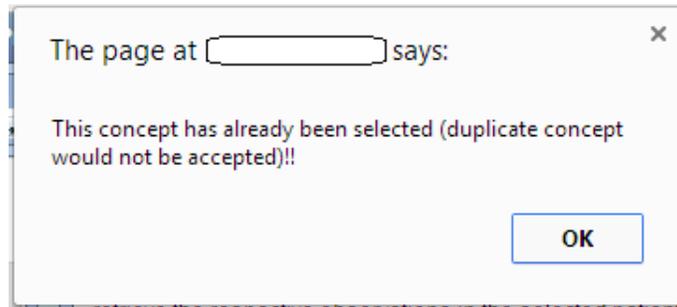


As implied in the pop-up dialog above, one can drop in many concepts:



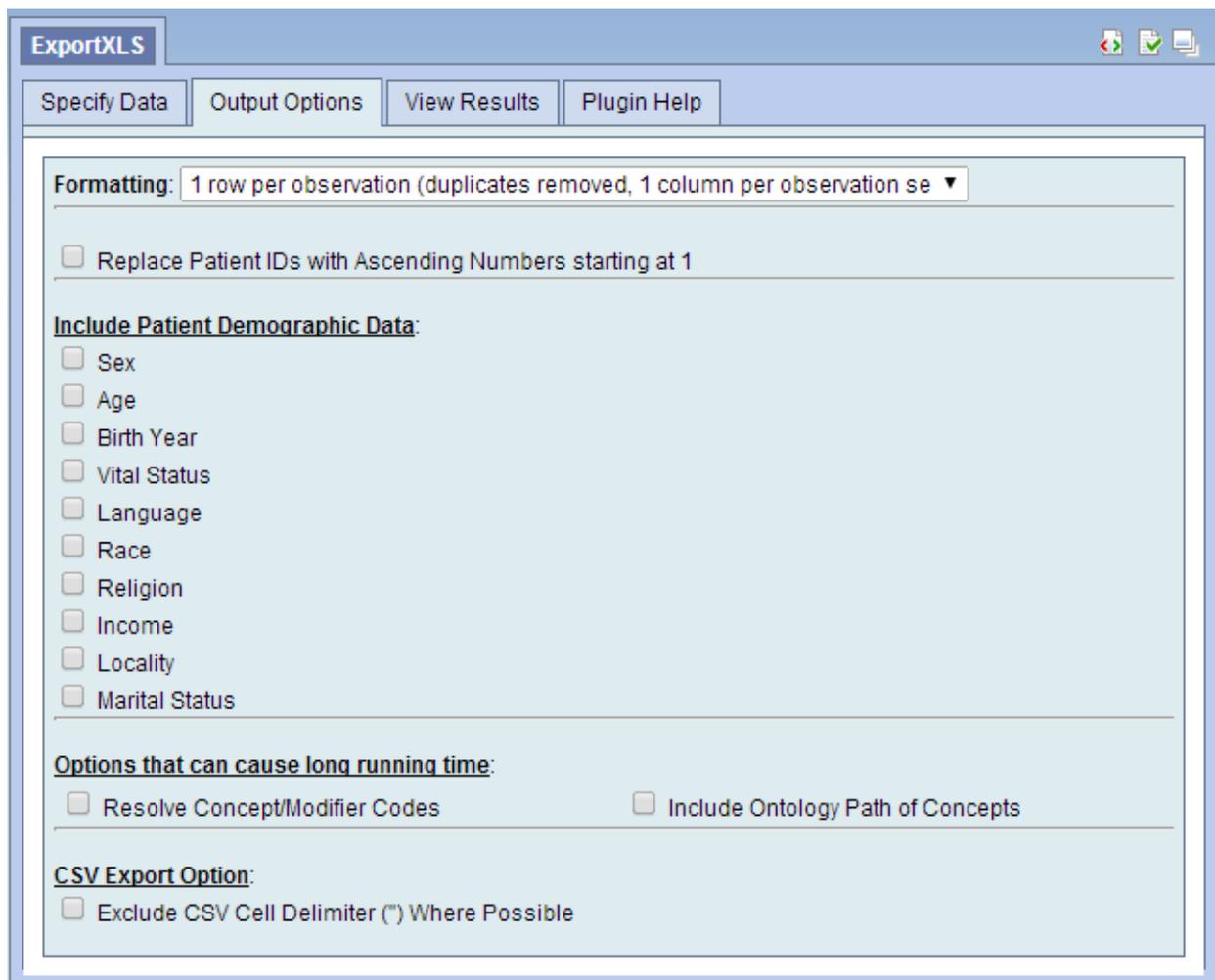
Notice that the instruction “Click a Concept to remove it from the list.” is displayed as soon as a concept is dropped into the box.

Also, a warning pop-up dialog will be displayed if a repeated concept is dropped in, similar to the following (from a Chrome browser that was used; note: IP address blotched out):



The “Output Options” Tab

Navigate to the "Output Options" tab to select the output format and options. There are four sections in this tab: the “Formatting” section, the “Patient Demographic Data” section, the “Options that can cause long running time” section, and the “CSV Export Option” section.

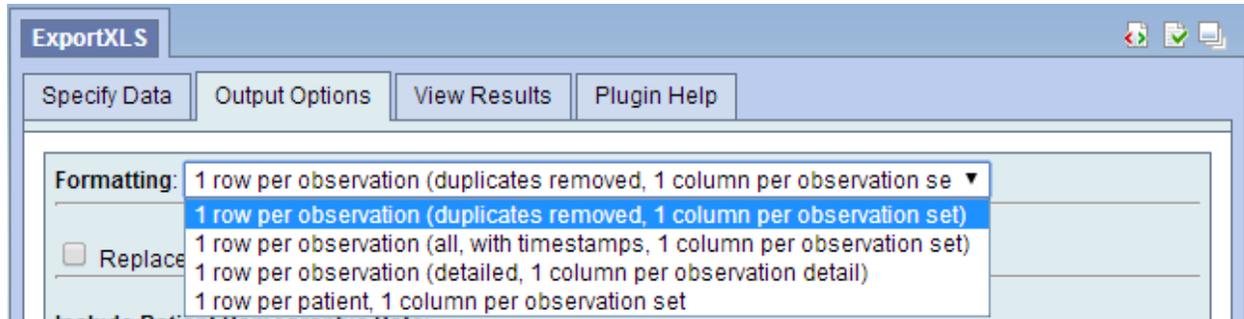


The “Formatting” Section

The “Formatting” section contains the “Formatting” selection list and the “Replace Patient IDs with Ascending Numbers starting at 1” checkbox.

The “Formatting” Selection List

There are four options in the “Formatting” selection list:



Their meanings are:

- 1 row per observation (duplicates removed, 1 column per observation set): A new row is created for each observation. All observation details (concept code, value, unit, ..., etc.) are written into one cell. One column is created for each concept that has been dragged onto the input box in step 1. **Attention:** Duplicate entries are removed! This format only returns a list of the **different** observations that were found.
- 1 row per observation (all, with timestamps, 1 column per observation set): Similar to the option above, **but** timestamps of the observations are tabulated as well. Therefore, duplicates are not possible and nothing is removed.
- 1 row per observation (detailed, 1 column per observation detail): This is the most detailed option. A new row is created for each observation and all observation details (concept code, value, unit, ..., etc.) are written to dedicated columns.
 - The 'Observation Value' column will contain the appropriate 'Observation BLOBs' provided that the 'valuetype_cd' is 'B' and the user has the appropriate access right ("DATA_DEID").
- 1 row per patient, 1 column per observation set: A new row is created for each patient. One column is created for each concept that has been dragged onto the input box. All observations of a patient are then written into one cell (with respect to the concept column). **Note:** This is the only output option where the first column starting number will match the specified value of 'Starting Patient'.

The “Replace Patient IDs with Ascending Numbers starting at 1” checkbox

If you want to replace the patient IDs with an ascending numbers (starting with 1), check the 'Replace Patient IDs with Ascending Numbers starting at 1' box. This works independently from other i2b2 anonymization / pseudonymization features and always replaces the patient IDs. Please consider: the numbers count for the current view of the table and are not related to real patient IDs. If the queried concepts are changed in a way that the resulting patient set is changed as well, the ascending numbers used to designate the patients of a previous query may not be the same in a new query!

The “Patient Demographic Data” Section

A number of demographic data items can be tabulated as well. Check the respective boxes to include these data items. Always selected only those data that are germane to the investigation at hand, to reduce cluttering in the resulting table, as well as faster turn-around time.

Please note that the list of demographic data options in your institution may differ from those shown in the illustration above, as there are built-in flexibilities to allow the administrator of each establishment to make available only those options that are relevant or compliant to its missions.

The “Options that can cause long running time” Section

There are two such “resource-intensive” options: 'Resolve Concept/Modifier Codes' and 'Include Ontology Path of Concepts'

The ‘Resolve Concept/Modifier Codes’ Option

If you want to resolve the concept and modifier codes (e.g. ICD codes) to see their denotation, check the box 'Resolve Concept/Modifier Codes'.

Warning: this will also require massive database querying and can therefore cause a very long running time or crash. It is a good idea to select this option only right before you are sure your dataset is complete and ready to be exported. As long as you are still varying your query, this option should be left unchecked.

The ‘Include Ontology Path of Concepts’ Option

If you want to include the ontology path of the concept codes, check the box 'Include Ontology Path of Concepts'.

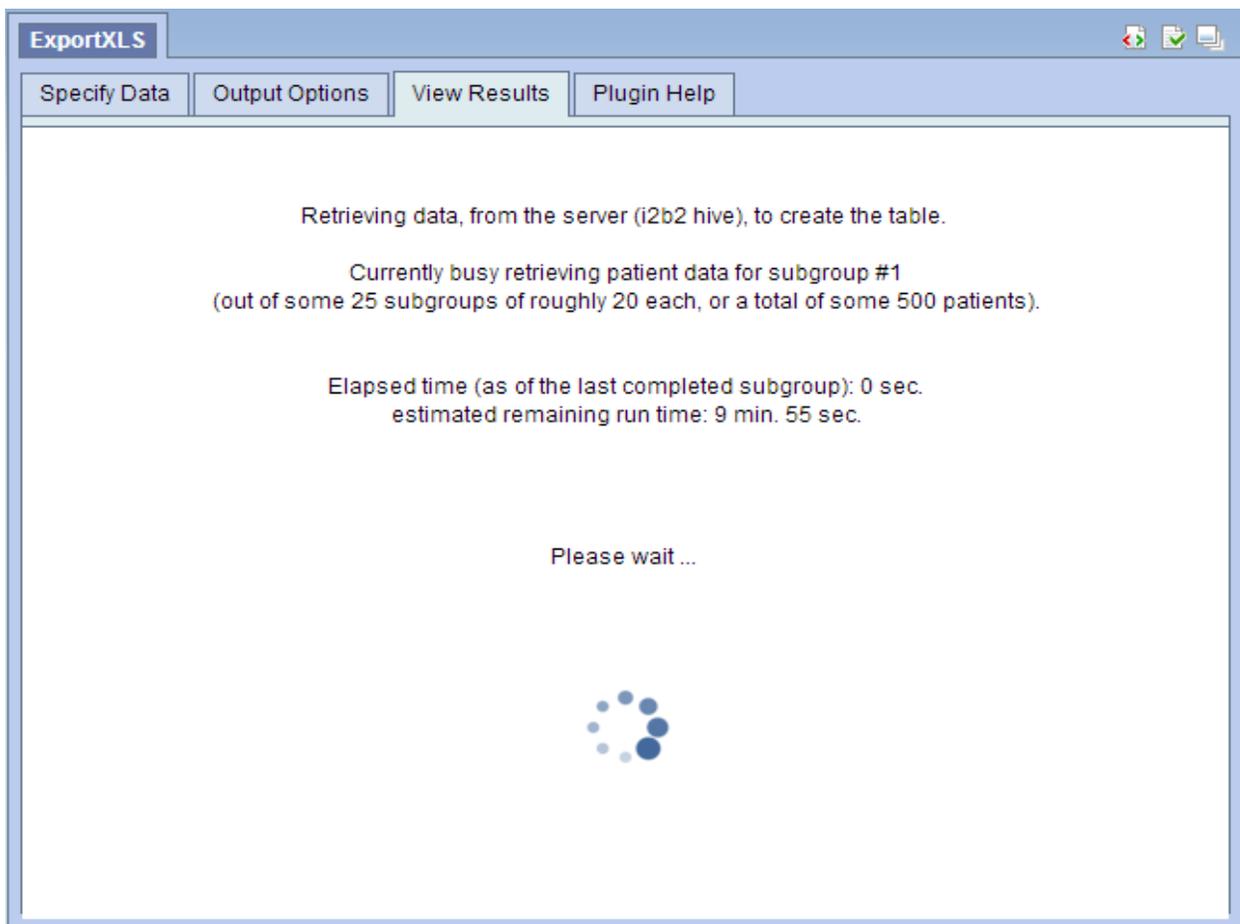
Warning: this also requires massive database querying and can therefore cause a very long running time or crash. However, if you check both 'Resolve Concept/Modifier Codes' and 'Include Ontology Path of Concepts' options, the running time will **not** double, since both options require the same ontology querying.

The “CSV Export Option” Section

The default export CSV file encloses all data cells in quotation marks ("). This is necessary to export special characters like line breaks within a cell. However, some statistics software may get confused with the quotation marks. In this case, check the 'Exclude CSV Cell Delimiter (") Where Possible' box to permit the writing of quotation marks where possible. Cells with special characters will still be enclosed in quotation marks, though... but it will be way less work to remove them manually than if it had to be done for every cell.

The “View results” Tab

Finally, select the "View Results" tab to view the progress and the resulting table of the observations.



Depending on the size of the Patient Set, the number and complexities of the concepts, and the various options specified, it could take quite a while for the i2b2 hive to complete the chore requested. However, while the data are being fetched, there will be constantly updated feedback displays to provide estimations as to how much data have been retrieved, how much are yet to be fetched, elapsed time, and remaining run time, etc.

Successful Run

At the end of a successful run, after all the requested data have been returned by the server (i2b2 hive), the resulting table will be displayed:

	Patient ID	Sex	Age in Years	Birth Year	Vital Status	Locality	Marital Status	Circulatory system
1	792584	M	69	1944	N	SPENCERMA		ICD9:414.00 = SECONDARY
2	1561286	M	67	1946	N	NORTHRIDGEMA	M	ICD9:414.02 = PRIMARY ICD9:414.01 = PRIMARY ICD9:401.1 = PRIMARY ICD9:414.00 = SECONDARY ICD9:401.1 = SECONDARY ICD9:401.9 = SECONDARY ICD9:414.01 = SECONDARY ICD9:414.00 = PRIMARY ICD9:411.1 = PRIMARY ICD9:414.2 = SECONDARY ICD9:414.05 =

Note that for the figure above, the “1 row per patient, 1 column per observation set” formatting option was selected.

If you want to export the table into a CSV file, click the **CSV Export** button. This initiates the download of a file onto your desktop. The file is named 'i2b2-export_patient-observations_[timestamp].csv'.

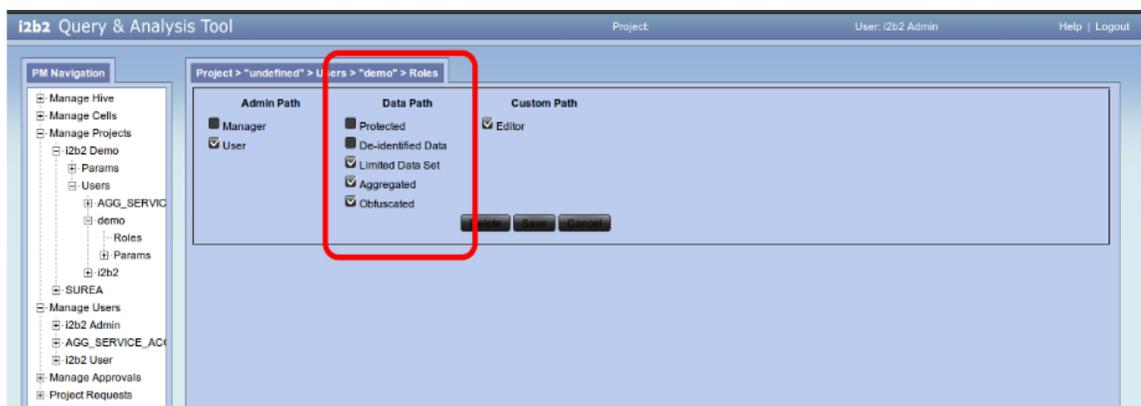
If you want to export the table into an HTML encoded Excel file, click the **HTML/XLS Export** button. This initiates the download of a file named 'i2b2-export_patient-observations_[timestamp].html.xls'. The file is HTML, but due to the suffix it can directly be opened with Excel. If you want to open it in a web browser, simply remove the '.xls' suffix.

Note: Excel may launch a warning message if the file is loaded. This message can be ignored/confirmed (click 'Yes').

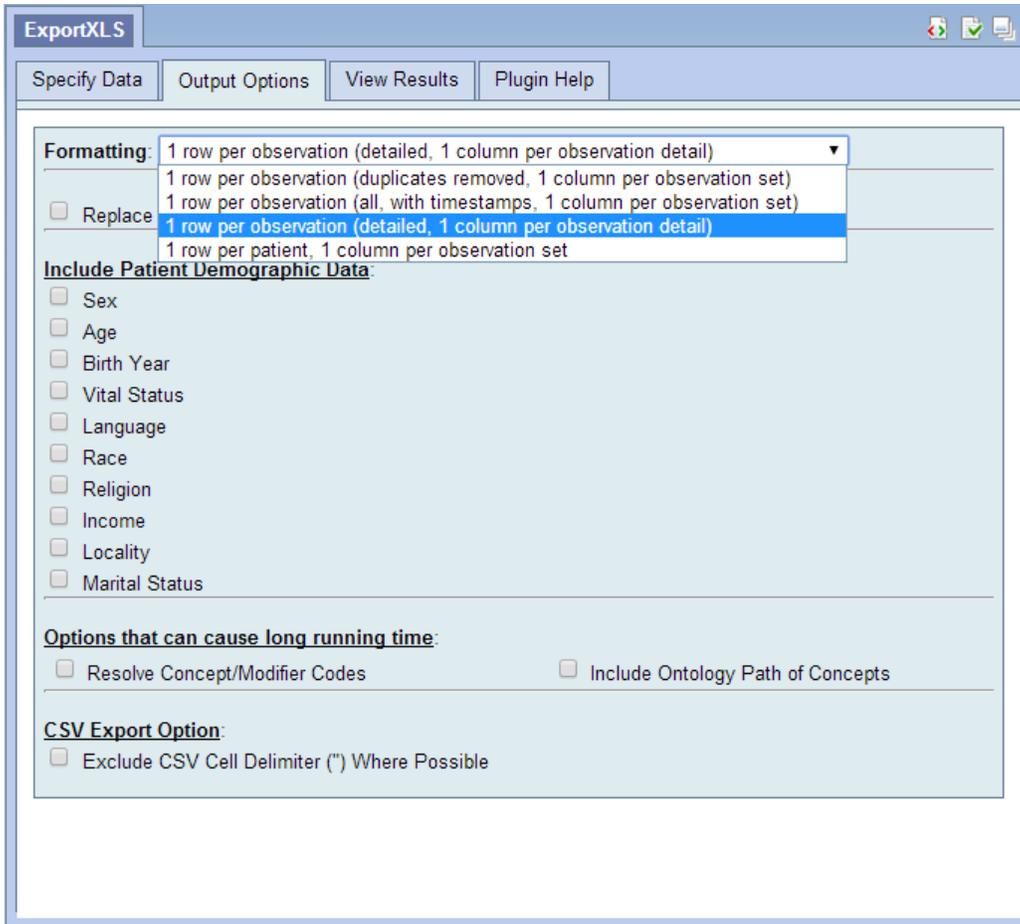
Displaying “Observation BLOBs”

The “Observation BLOBs” are only available when the following conditions are met:

1. User has been grant the “DATA_DEID” access (by the i2b2 Administrator)



2. User has selected the “1 row per observation (detailed, 1 column per observation detail)” Formatting option:



3. The corresponding observation “valuetypce_cd” is “B”.
4. Observation BLOBs are actually available in the database.

The following is an example of “Observation BLOBs” displayed in the “Observation Value” column of the resulting table:

Patient ID	Visit ID	Provider ID	Timestamp Start	Timestamp End	Instance Number	Observation Set	Observation Concept Code	Observation Value	Observation Unit	Observation Modifier
1	1000000001	471260	LCS-I2B2:0000109060	1997-03-25	1997-03-25	1	Reports	LCS-I2B2:FH	Positive for diabetes, hypertension and kidney disease	
2	1000000001	471260	LCS-I2B2:0000109060	1997-03-25	1997-03-25	1	Reports	LCS-I2B2:SH	Married and lives with husband and one son. Patient denies tobacco, alcohol or illegal substance abuse.	

Note that this enhancement brings the access of “Observation BLOBs” (that was only available through the Workbench) to the Webclient.

In the case that the user does not have the “DATA_DEID” access, then only “[protected]” would be displayed instead:

ExportXLS

Specify Data | Output Options | View Results | Plugin Help

Click on one of the buttons on the right to download the following table in the appropriate format. CSV Export HTML/XLS Export

Patient Information (for patients 1 - 52)
of Patient Set 'Female@16:36:24 [3-11-2014] [demo] [PATIENTSET_167]'

	Patient ID	Visit ID	Provider ID	Timestamp Start	Timestamp End	Instance Number	Observation Set	Observation Concept Code	Observation Value	Observation Unit	Observation Modifier
3	1000000001	482694	LCS-I2B2:D000109060	2009-06-01	2009-06-01	1	Reports	LCS-I2B2:CT_RPT	[protected]		
4	1000000001	482694	LCS-I2B2:D000109060	2009-06-01	2009-06-01	1	Reports	LCS-I2B2:CT_RPT_DID	[protected]		
5	1000000002	476659	LCS-I2B2:D000109060	2004-06-03	2004-06-03	1	Reports	LCS-I2B2:CT_RPT	[protected]		
6	1000000002	476659	LCS-I2B2:D000109060	2004-06-03	2004-06-03	1	Reports	LCS-I2B2:CT_RPT_DID	[protected]		
7	1000000005	475950	LCS-I2B2:D000109062	2003-09-09	2003-09-09	1	Reports	LCS-I2B2:MRI_RPT	[protected]		
8	1000000005	475950	LCS-I2B2:D000109062	2003-09-09	2003-09-09	1	Reports	LCS-I2B2:MRI_RPT_DID	[protected]		
9	1000000006	473265	LCS-I2B2:D000109061	2000-08-10	2000-08-10	1	Reports	LCS-I2B2:XR_RPT	[protected]		
10	1000000006	473265	LCS-I2B2:D000109061	2000-08-10	2000-08-10	1	Reports	LCS-I2B2:XR_RPT_DID	[protected]		
11	1000000009	473136	LCS-I2B2:D000109061	2000-06-14	2000-06-14	1	Reports	LCS-I2B2:XR_RPT	[protected]		
12	1000000009	473136	LCS-I2B2:D000109061	2000-06-14	2000-06-14	1	Reports	LCS-I2B2:XR_RPT_DID	[protected]		

Note that several new columns have been added:

ExportXLS

Specify Data | Output Options | View Results | Plugin Help

Click on one of the buttons on the right to download the following table in the appropriate format. CSV Export HTML/XLS Export

Patient Information (for patients 1 - 100)
of Patient Set 'Ischemi-Male@13:25:22 [11-21-2013] [demo] [PATIENTSET_281]'

	Patient ID	Visit ID	Provider ID	Timestamp Start	Timestamp End	Instance Number	Observation Set	Observation Concept Code	Observation Type	Observation Value	Observation Operator	Observation Unit	Confidence Number	Observation Modifier
1	147494	7699218	undefined	2008-02-07	2013-11-18	1	Circulatory system	ICD9:414.01	T	PRIMARY			0	
2	147494	8489912	undefined	2008-02-07	2013-11-18	1	Circulatory system	ICD9:414.01	T	PRIMARY			0	
3	147494	8992543	undefined	2008-02-07	2013-11-18	1	Circulatory system	ICD9:425.4	T	SECONDARY			0	
4	147494	8296485	undefined	2008-02-07	2013-11-18	1	Circulatory system	ICD9:425.4	T	SECONDARY			0	
5	254	7457405	undefined	2006-08-	2013-11-	1	Circulatory	ICD9:401.0	T	SECONDARY			0	

The “Observation Type” column indicate whether “Observation Value” entries are “numeric”, “text”, or “BLOB” for the user.

The “Confidence Number” column will provide the measure of the confidence of the observation.

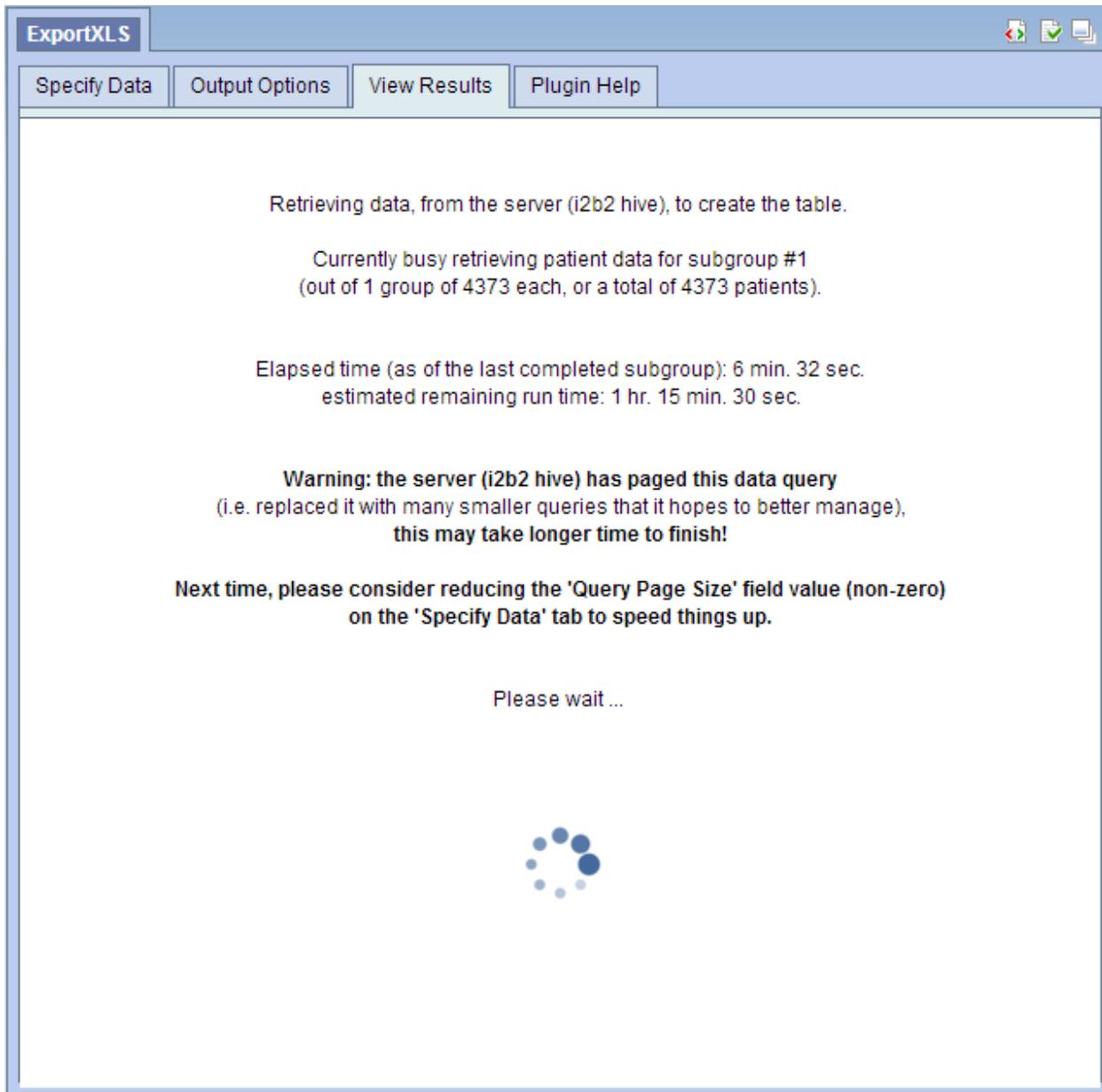
The “Observation Operator” column display (when the “valuetype_cd” = “N”) the qualifier (“=”, “>”, “<”, “=>”, “<=”, or “!=”) for the numerical “Observation Value” (e.g. “>” means the “Observation

Value" is "greater than" the one shown (refer to the i2b2 Data Repository (CRC) Cell Design Document for further details).

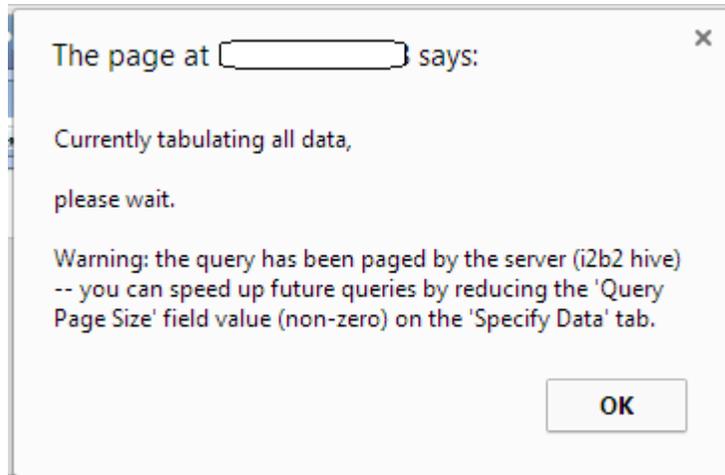
Warning About The i2b2 Hive Paging

When a user specifies very complex Concepts against very large Patient Set; the i2b2-hive may attempt to page that large query (or separate it into smaller sub-groups of queries, and then combine the results of these smaller queries into one set of data to be returned). However, sometimes it may lose track of its pages and cause an error resulting in an apparent hang in this plugin (i.e this plugin appeared to be never finishing).

When it is detected that the i2b2-hive has paged the request, this plugin will display a warning (along with its estimates) message to urge the user to try options that would avoid future pagings from occurring, like the following:



If the paging did not result in error, then a similar warning would also be displayed in the “all data retrieved” pop-up dialog, similar to the following (from a Chrome browser that was used) to again remind the user to try to avoid this same predicament in future runs (note: IP address blotched out):

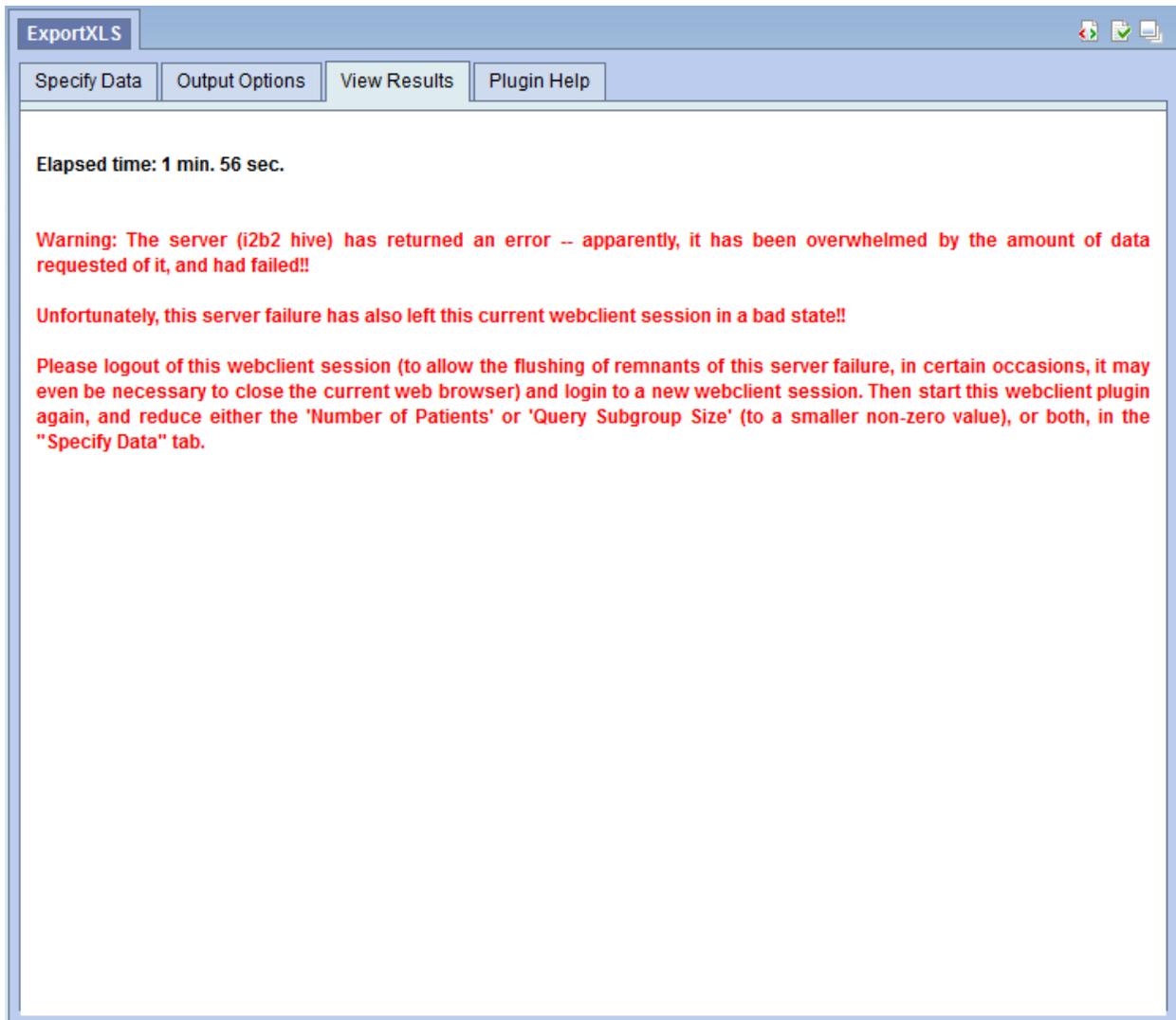


Unsuccessful Run

In previous versions of this plugin, because of its popularity, user has been known to specify more and more complex Concepts against large Patient Set; often time resulting in hang (i.e this plugin appeared to be never finishing).

The cause of such situation has traced to the i2b2-hive somehow losing track of its pages when it attempted to page the large query (or separate it into smaller sub-groups of queries, and then combine the results of these smaller queries into one set of data to be returned).

Despite of all the additional options introduced, it is still possible for one to have selected options that sensitize this problem of the i2b2-hive's. When that occurs, then the following display will result:



Caveats

Please note the following:

- Since it is possible for the server (i2b2 hive) to take a long time to provide all the data requested, this plugin has been enhanced with occasional updates of 'Elapsed time' and 'estimated remaining run time', etc. displays. These displayed values are best rough estimates based on occasional data coming back from the server. In the case of zero 'Query Subgroup Size' selected (i.e. no query-subgrouping), then there would be no data coming back from the server until either of the following situations:
 - The whole, single, query is done.
 - The server itself started paging (i.e. query-subgrouping), and the result of the 1st of such paged subqueries just arrived.
 - The server got overwhelmed by all the excessive data and failed, returning an 'error'.

In any of these cases, the updates of 'Elapsed time' and 'estimated remaining run time' will be quite infrequent, as updates from the server (i2b2 hive) will most likely take a long time, and far and few in-between, if any.

- This plugin also tries to provide updates on the current subgroup the server (i2b2 hive) is fetching. However, in the cases when the server (i2b2 hive) itself started paging (i.e. query-subgrouping), then the current subgroup # being fetched may exceed the original stated total number of subgroups.
- The '1 row per patient, 1 column per observation set' option of tabulation is the only output option where the first column starting number will match the specified value of 'Starting Patient'.
- The '1 row per observation (detailed, 1 column per observation detail)' option of tabulation is the only output option where the 'Observation Value' column will contain the appropriate 'Observation BLOBs' (if the 'valuetype_cd' is 'B' and the user has the "DATA_DEID" access right).

Browser Compatibilities

- This plugin has been successfully tested with the following browsers: Firefox (v18.0.2, v28.0), Internet Explorer 32/64 (v9.0.13, v10.0, v11.0 - *there seems to be a problem with scrolling bars in small windows, though*), Chrome (v24.0, v33.0), Safari (v5.1.7) and Opera (v18.0, v20.0).
- The i2b2 web client itself does not work with Opera (v12.11, v12.16).

i2b2 version compatibilities

This current version is compatible with i2b2 v.1.3-1.7.

Version History

1.0	Initial release, for i2b2 v1.5. By Mauro Bucalo, Universita' di Pavia, Italy.	06/11
1.6	Updated for i2b2 v1.6. By Mauro Bucalo, Universita' di Pavia, Italy.	01/12
2.0	Updated for i2b2 v1.3-1.6, Internet Explorer compatibilities; added flexible column exclusion, table title & caption, leading row count column, and other enhancements. By Wayne Chan, University of Massachusetts Medical School, Worcester, USA; and Mauro Bucalo, Universita' di Pavia, Italy.	02/12
3.0	Major upgrade; many new features (CSV export, detailed tabulating of all observations, formatting options, paged queries...), heavily re-engineered. By Axel Newe, Friedrich Alexander University Erlangen-Nuremberg, Germany.	02/13

3.2	Added options of using a subset of a large patient set; fixed blank 'Birth year' bug; added 'locality' column per request; added more user-friendly feedback (estimated run-time, etc.), help & hint; added server failure termination & notification. By Wayne Chan, University of Massachusetts Medical School, Worcester, USA.	11/13
3.3	Added support of displaying observation blobs (for users with appropriate access) and added several useful columns to the output option of '1 row per observation (detailed, 1 column per observation detail)', etc. By Wayne Chan, University of Massachusetts Medical School, Worcester, USA; and PARIS Nicolas, Hôpital Ambroise Paré, Boulogne, France.	03/14

Terms of Use

This plugin is published under the [GNU GPL v3 license](#).

This software is distributed as a plugin to the i2b2 v1.3-1.7 webclient framework, and may be used free of charge. It is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.