Publishing Microdata to <odesi> Using Nesstar Publisher 4.X

Quick Reference Guide

August 21, 2012

Alexandra Cooper, Queen's University

TABLE OF CONTENTS

List of Revisions 2

Importing from Nesstar Publisher 3.x 3

Preparing the Dataset 4

TIPS for marking up files 7

Creating / Editing the Dataset 8

1) Importing Dataset 8

2) Resequence Variables 8

3) Declare Weights for entire file 9

4) Add weight variable to individual variables 9

5) Marking up the dataset 10

6) Validate 14

7) Save file 14

8) Exporting Metadata (.xml file) 14

9) Publishing Data and Metadata 15

10) Republish 15

Re-importing the data file into Nesstar 15

LIST OF REVISIONS

June 8, 2012

- Guide created (Queen's)

July 24, 2012

- added Importing from Nesstar 3.x
  (Revisions made by Queen's)

July 30, 2012

- added to 6) Marking Up the Dataset note about tags to be edited when using XML file from other source
  (Revisions made by Queen's)

IMPORTING FROM NESSTAR PUBLISHER 3.X

If you have a Nesstar 3.x file (.NSDStat) it is very easy to import it into Nesstar 4.0.

1) **Importing .NSDStat file**

   Some files can take a while to import – be patient

   a) Open Nesstar Publisher 4.0
   b) Click on My Projects
   (if this is not highlighted, then the Import Study option will not be available)
   c) Click on **File → Import Study…**
   d) Select the .NSDStat from the drop down box → Double click on file name
   e) File will open in the right hand window
   f) Save the file→ **File → Save**
   g) If necessary, make any edits and save new edits
   h) Publish

Note: most .NSDStat files will likely have ID# with underscores (_). These need to be changed to dashes (-).

PREPARING THE DATASET

Before marking up the file (entering the metadata into Nesstar Publisher), always check to ensure that the file is ready for public use. You will need to check for the following:

- there is an id variable;
- all the variable labels are correct and unique (avoid lengthy labels/repeats that read as a literal question);
- all the value labels are correct;
- all the missing values have been declared;
- all the recoded variables are correct;
- any alphanumeric variables are truly alphanumeric; and
- make sure the data are not weighted.

This guide assumes you will perform these checks using SPSS. However, sometimes quick fixes can be made in Nesstar. Generally, you will have either an SPSS system file (.sav format – obtained from <odesi> or Statistics Canada Nesstar), or you will have a combination of an ASCII text file (.txt or .dat) and a syntax file (.sps – obtained from Statistics Canada FTP site). The following will describe how to prepare the dataset in each case.

1) Working with an ".sps" syntax file (creating an SPSS system file .sav)

Open the syntax file (.sps file) by double-clicking on it. The .sps file is the program file which converts the raw data file (ASCII text file) to an SPSS system file (.sav) which is then imported into Nesstar.

a) Modify the SPSS command code

See below for instructions on editing the syntax file.

Tips on editing SPSS command code:

- all commands end with a period.
- all commands start in the first column; subsequent lines in the same command are indented by 1-2 columns.
- all lines of text are in quotation marks
- LRECL (Logical Record Length) can be determined by looking at the last position of the last variable on the raw data file. Or, the last number at the end of the DATA LIST FILE.

b) Run the file

SPSS will run the syntax file's code against the raw data file referred to.

i) Save the file

ii) From the Menu bar, use the menu command: Run → All to run the code. When it is done, "SPSS processor is ready" will appear at bottom of screen.

c) Check results.

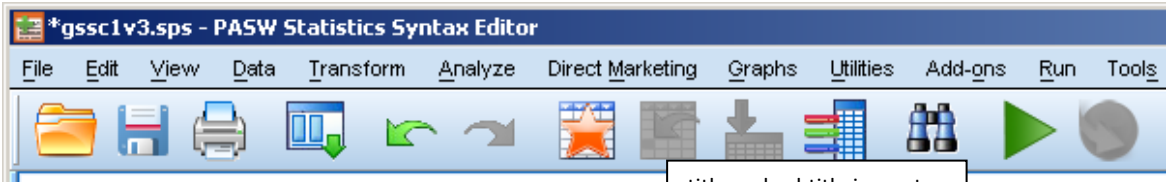
i) Check that the SPSS System File (.sav) was saved to the disk location you specified at the bottom of your SPSS program (.sav file will have been added to working directory / folder)

ii) From within the data editor screen (the one that looks like a spreadsheet), use the menu command: **Analyze → Descriptive Statistics → Frequencies**

iii) Choose an appropriate variable (e.g. Province, or some other categorized variable)

iv) Run the analysis and compare your results (e.g. number of cases) with the number of cases indicated in the "Readme" file or Data Dictionary for the survey

v) The .sav file can now be imported into Nesstar Publisher




```
TITLE "GENERAL SOCIAL SURVEY CYCLE 1, 1985"
SUBTITLE "HEALTH AND SOCIAL SUPPORT"
SET LENGTH=NONE WIDTH=80.
FILE HANDLE GSS/NAME='C:\data\gssc0185.dat'/LRECL=559.
```

title and subtitle in quotes

sets page width

identifies file, gives record length

'nickname' of file same as DATA LIST FILE name

full path to where raw data is located

Logical Record Length (LRECL)

```
DATA LIST FILE=GSS/
SEQNO        1 - 5
GENHLTH      6 - 6
BPCHECK      7 - 7
```

numbers identify positions of variables in raw data file

.....

```
ICIMACT1     22 - 22
DAYSMISS     23 - 24   (A)
CUTDOWN      25 - 25
CUTALLDY     26 - 27
ICIMACT2     28 - 28
DAYSNTAB     29 - 30   (A)
SEEDOC       31 - 31
```

"**(A)**" – indicates that this variable is alphanumeric;
It's value label will be letters not numbers.
If the data are numbers, this has to be changed.
See Appendix C for instructions on how to change alphanumeric variables to numeric variables.

.....

```
WEIGHTI      91 - 93
WEIGHTM      94 - 97   (1)
OVRUNDR      98 - 98
EXERCISE     99 - 99
TYPEXER1     100 - 100
```

"**(1)**" – indicates the number of decimal places to be read in; in this case, this variable should be interpreted as having 1 decimal place.

.....

```
GCHILD       556 - 556
RELATIV      557 - 557
EXFARE       558 - 559.
```

a period (" . ") signifies the end of this command

this number is the LRECL – it is the last position of the raw data file

```
VARIABLE LABELS
SEQNO      "SEQUENCE NUMBER"
GENHLTH    "DESCRIBE YOUR STATE OF HEALTH"
BPCHECK    "TIME SINCE LAST BLOOD PRESSURE CHECK"
GCHILD     "HOUSEHOLD INCLUDES RESPOND GR-CHILDREN"
RELATIV    "HOUSEHOLD INCLUDES RESPONDENTS RELATIVES"
EXFARE     "EXTENDED FAMILY SUMMARY VARIABLE".
```

Make most changes in this section.
Clean up variables if needed:
- check spelling
- spell out words in full (i.e cigs – cigarettes)

```
VALUE LABELS
GENHLTH
    1  "EXCELLENT"
    2  "GOOD"
    3  "FAIR"
    4  "POOR"
    9  "NOT STATED"
   /
BPCHECK
    1  "6 MONTHS OR LESS"
    2  "7-12 MONTHS"
    3  "13-24 MONTHS"
    4  "2+ YEARS"
    5  "NEVER"
    6  "DO NOT KNOW"
    8  "TIME UNKNOWN"
    9  "NOT STATED"
    .
```

Generally, don't change things in this section.

```
MISSING VALUES
GENHLTH  (9)
BPCHECK  (6 9)
BPHIGH  (0 3 9)
BPTREAT  (0 3 9).
```

Always check to make sure all missing values have been declared. All <odesi> files should have missing values declared.

```
SAVE OUTFILE="C:\data\gss1.sav".
EXECUTE.
```

Include the full path where the resulting SPSS sytem file (.sav) is to be saved. Give the file a meaningful name (survey acronym and year).

Changing Alphanumeric Variables to Numeric Variables in the SPSS Syntax File

In some instances, variables will be coded as alphanumeric when they are really numeric variables. In these cases, the variables need to be changed to numeric in the SPSS syntax file.

Alphanumeric variables are noted by the '*(A)*' beside the variable in the Data List. To check if the variable is truly an alphanumeric variable, scroll down to the Value Labels and look at the values for the variables. If the value of a variable is listed in quotes (*'a' "Excellent"* or *'1' "6 months"*) then the variable is coded as alphanumeric. If the value is a letter, then it is fine; if the value is a number, then it needs to be changed.

To change a variable from alphanumeric to numeric:

- **i)** remove the '*(A)*' from the Data List
- **ii)** remove the quotes from around the value in the Value Labels list
- **iii)** save file and run syntax file with data file to create new .sav file
- **iv)** import the new .sav file into Publisher

For more information on this, see Appendix C in **Publishing to <odesi> Using Nesstar Publisher 4.0 (using DDI2.x)**.

2) **Working with an SPSS system file (.sav) file.**
If you are starting directly with a .sav file, you do not need to go through the process that was described in part 1) above.
- Open the .sav file in SPSS
- Go through the checklist listed at the beginning of this section.
- Make any needed changes.
- Save the file.
- The .sav file can now be imported into Nesstar Publisher.

TIPS FOR MARKING UP FILES

- For *Document Description* and *Study Description* the correct title is necessary, including the year and study # (if applicable).
- In the left-hand side tree navigation, click on '+' to expand a section; click on the folder name to see all fields in that section; click on the field name to edit it.
- When you are filling in the information on the right-hand side and you want to add more entries, click on the '+' sign on the right-hand side. This can be done as often as necessary. It will ensure that the entry is on a different line when it is published. If there is no '+' sign, then it is not possible to add more entries in that particular field.
- Likewise, you can click on the '–' sign to take away an entry you don't want.
- If you want to re-order your entries, put the cursor over the entry you want moved and click once so there is a box around it, and then click on the up or down arrow on the right-hand side.
- If there is a Globe icon on the right-hand side, clicking on it will enable you to put in a URI (external link) that is relevant to that particular field.
- Correct spelling and grammar are of the utmost importance.
- It may seem like you are repeating yourself when you are entering metadata in different sections and that is because you are! Remember that the different fields you are entering information into will appear in different parts of the finished document and must be complete on their own.

☆ *Always remember to SAVE the file and export the .xml file regularly!*

CREATING / EDITING THE DATASET

1) Importing Dataset

Some files can take a while to import – be patient

a) Open Nesstar Publisher

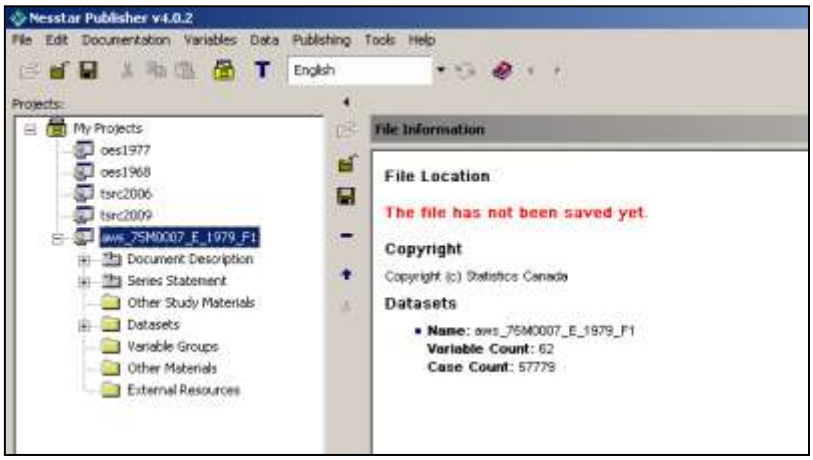
b) Ensure that you're using an appropriate template. Nesstar always uses the last template used.

c) Click on My Projects
(if this is not highlighted, then the Import Study option will not be available)

d) Click on **File → Import Study…**
(to import more than one file, click on **File → Import Multiple Studies…**)

e) Select the file type (.sav) from the drop down box → Double click on file name
(to choose more than one file, hold CTRL or Shift and click on file names)

When data is imported, you will see this screen. To add/edit metadata, click on the section names on the left. To see the data, click on the '+' beside 'Datasets', then on 'Variables'.



f) Save and rename the file

When the dataset is first imported, Nesstar uses the name of the dataset as the filename.
File → Save As… to save and rename the file
The file will save with a .Nesstar extension.

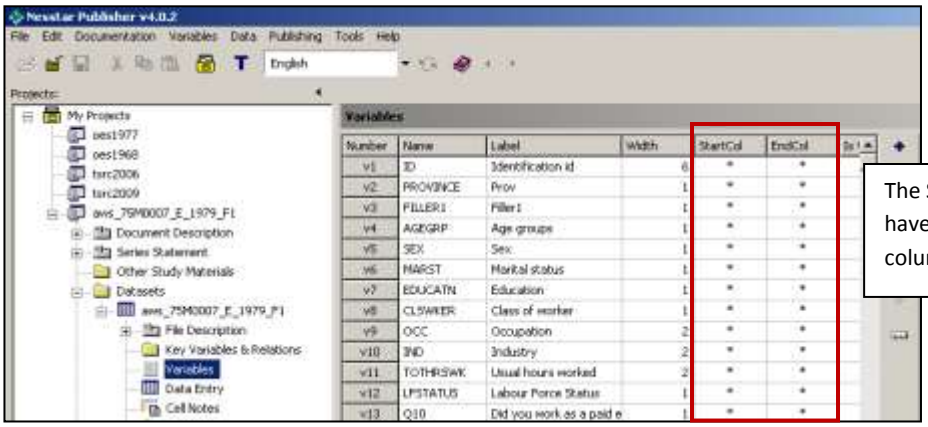
2) Resequence Variables

This is very important to do and must be done for each dataset. It recalculates the start and end positions for each of the variables. If it is not done, then the user will get an error when trying to export to spss, sas or ASCII.

a) Click on '+' beside Datasets → click on Variables

b) Variables → Resequence

c) *Resequence* dialog box – "Do you want to recalculate the widths? Recalculating the widths are recommended if you plan to export as Fixed Format Text, SAS or SPSS syntax." → Yes



The StartCol and EndCol will have an asterisk until the columns are resequenced

3) Declare Weights for entire file

State which variable is the weight variable (repeat for each weight variable). This applies the weight to the entire file; to add weight to individual variables, see Add weight variable to individual variables below.

a) Choose weight variable (user guide will list weight variable(s))

b) In lower right-hand corner, under Variable Information, check Is Weight Variable (do this for each weight variable)

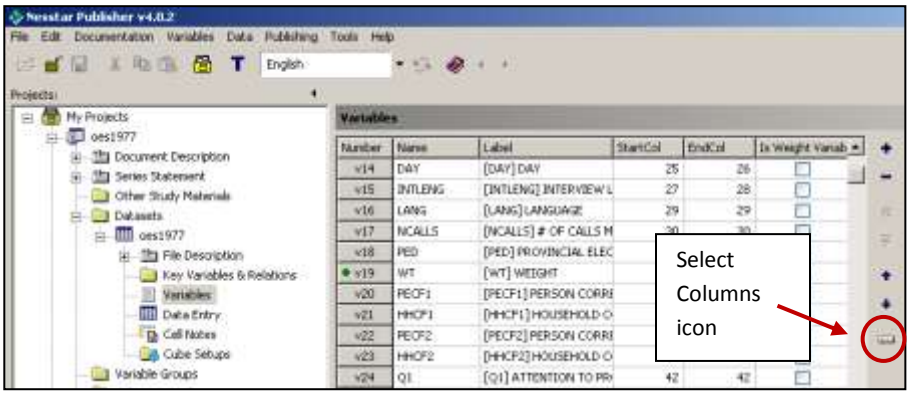
OR

Turn on **Is Weight Variable** checkbox

c) Click on Select Columns icon

d) Customize the variable list dialog box opens

e) Click Is Weight Variable → OK

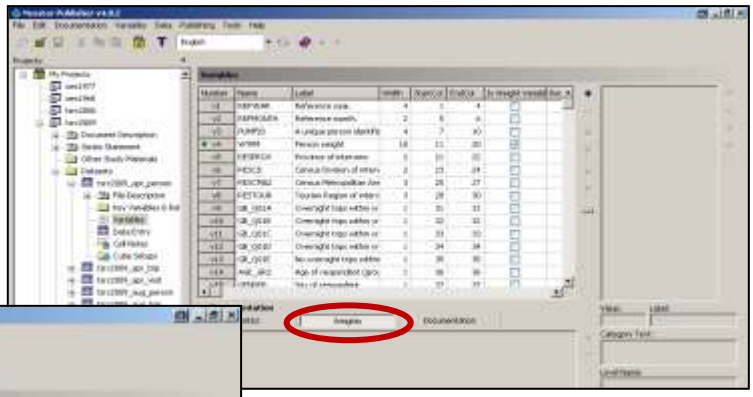


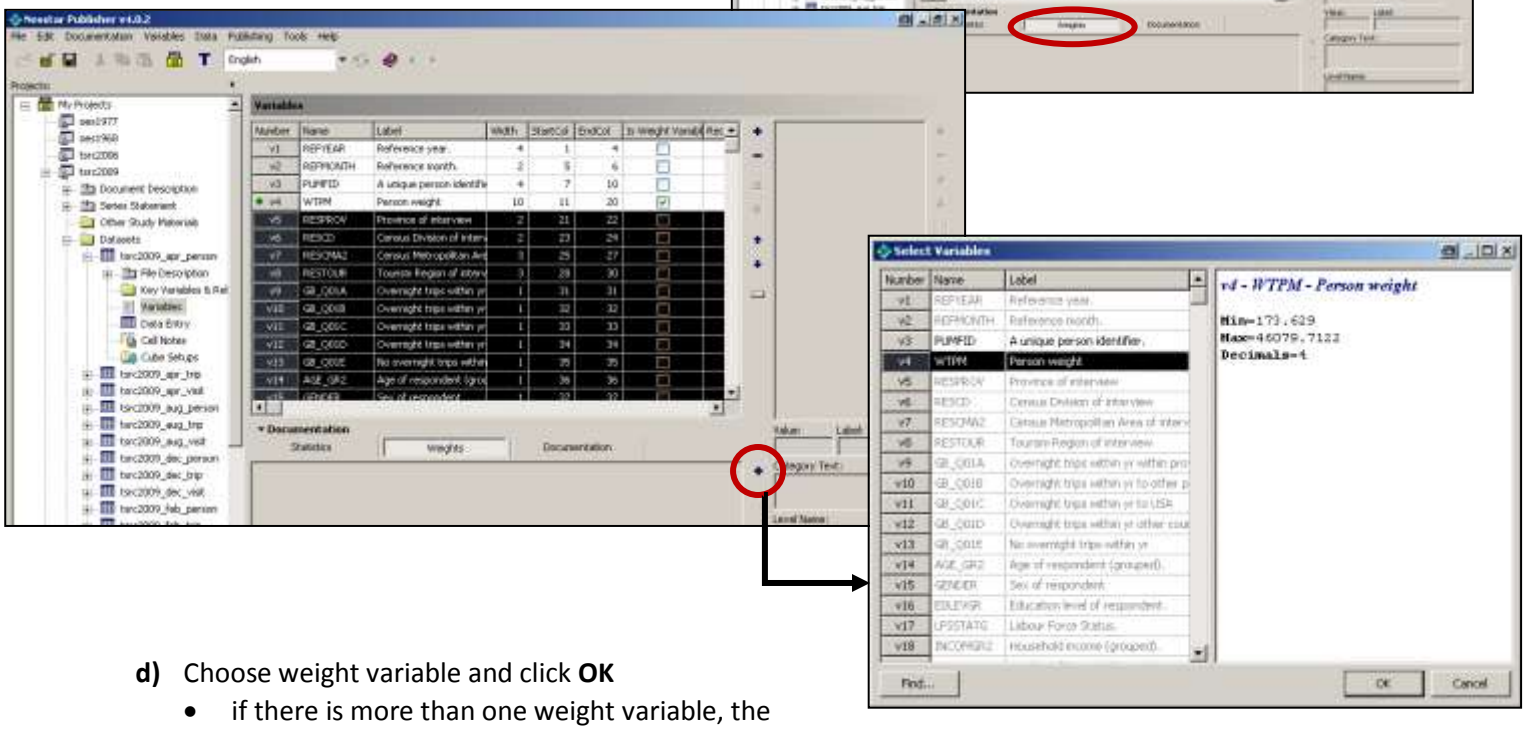
4) Add weight variable to individual variables

a) Click on Weights tab in the middle of the screen

b) Highlight variables that weight is to be added to

c) Click on '+'; the *Select Variable* dialog box opens

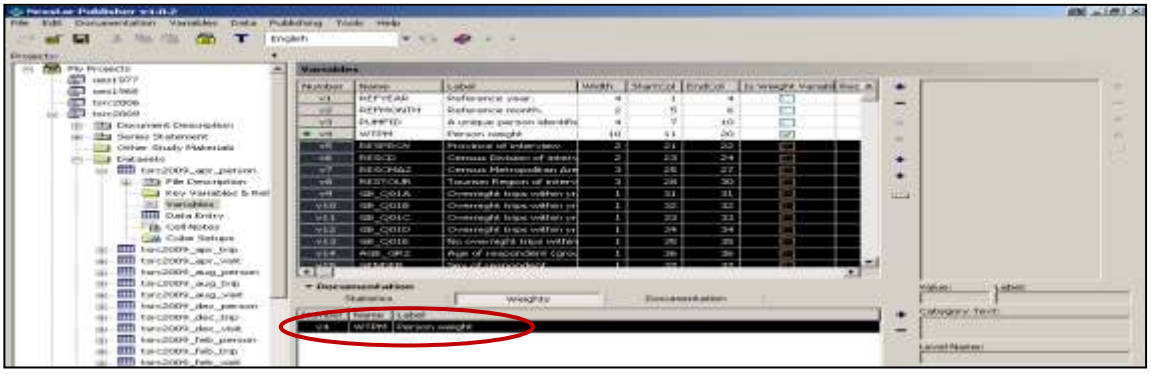




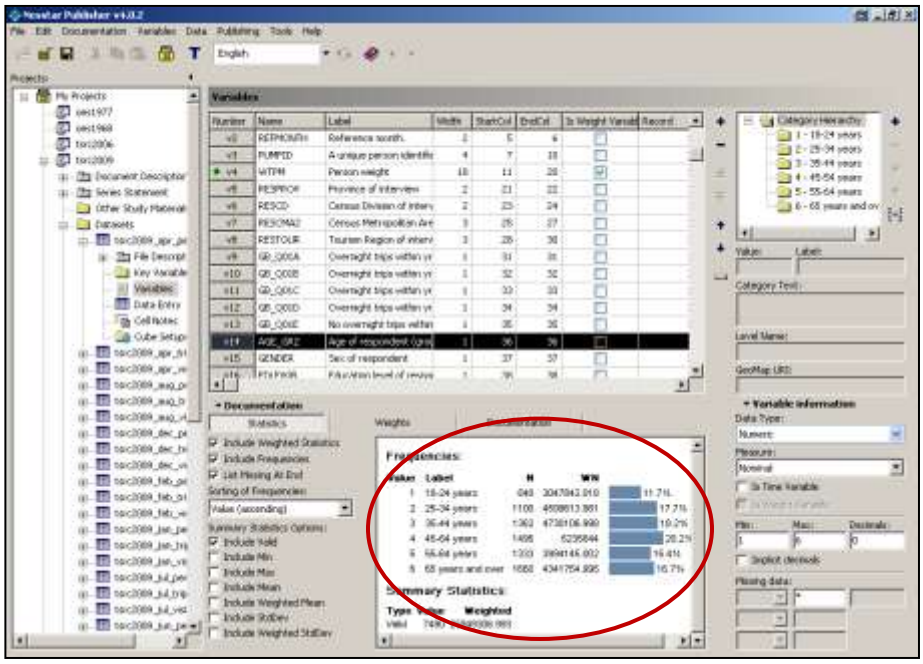
d) Choose weight variable and click **OK**

- if there is more than one weight variable, the proper weight variable must be applied to the appropriate variable. The User Guide will usually have this information.
- if there is more than one weight variable, and each one can be applied to any of the variables (for example, an individual weight, a household weight), then **<u>do not</u>** do this step.

e) Weight variable is now listed under the **Weights** tab.



f) Click on **Statistics** tab, choose variable that weight has been added to.
The **Preview** window will now show the variable with both weight and non-weighted values.



5) Marking up the dataset

Below is the list of sections in which you will be entering information. Each section contains a number of fields that must be filled out. For more detailed information and examples, please refer to the Best Practices Document (http://spotdocs.scholarsportal.info/display/odesi/best+practices+-+english).

After entering the metadata, the last steps in marking up the dataset are validating the metadata, saving the .Nesstar file and exporting the .xml file.

If you have existing metadata in XML format, you can import it rather than starting from scratch.

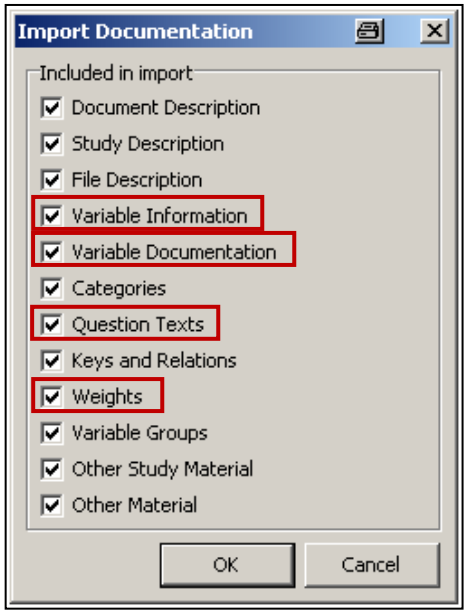
Importing metadata from other source:

After the dataset has been imported:

Documentation → Import → Import from DDI… (.xml file) **OR**
Documentation → Import → Import from Study… (.Nesstar file)

These two options are essentially the same. One uses the .xml file the other, the .Nesstar file.

i) *Import Documentation* dialog box
ii) Check the sections you want to import
iii) Click **OK**
iv) Find the .xml file
v) Click on **Open** and the metadata will be added to your file



Variable Information – imports the following:

- variable label
- location information (start column, end column, width, record number)
- decimals for numeric variables
- measure definition for numeric variables
- missing definitions for numeric and fixed string variables
- min and max for numeric and data variables are reset
- statistics options are reset for numeric variables
- universe and notes for variables

Variable Documentation – imports all the documentation for a variable, except for question text and variable labels

Question Text – imports the literal, pre- and post-questions and interviewers instructions; universe and notes are imported with Variable Information

Weights – if weights are declared in the source file, then they will be copied when imported (for both the file and individual variables)

NOTE: when importing, the file imported will override any previous default fields from the template.

NOTE: for a list of tags that must be edited when using metadata from another source, see section **7)f) Tags that must be edited when XML file is imported** in **Publishing Microdata to <odesi> Using Nesstar Publisher 4.x**.

Sections to be Marked Up

a) Document Description

Describes the metadata, that is, the document which you are presently putting together.
Click on '+' beside **Document Description** to open section. Click on field name to edit.

b) Series Statement

Describes the study

NOTE: for Beyond 2020, Excel files, and any other aggregate data only:

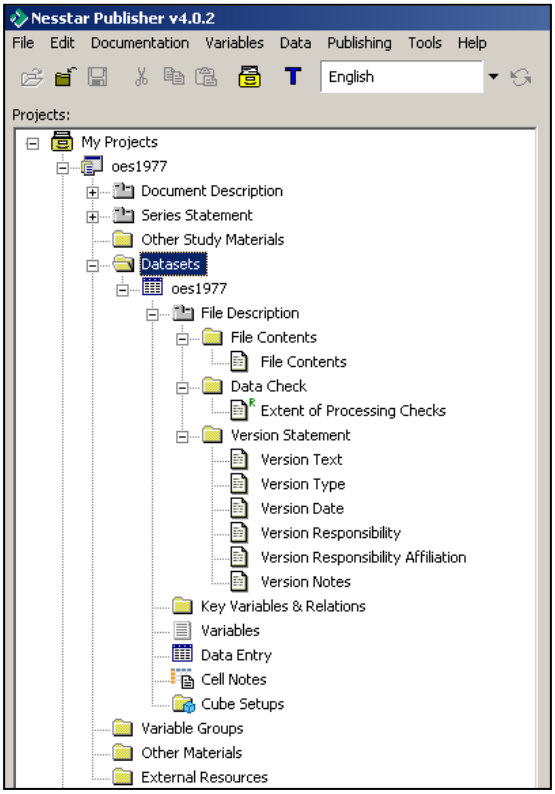
- 2.4.1.1 Location of Data Collection - this is the only time this field would be used
- Enter the link to the aggregate data file on the <odesi> ftp server:
  http://odesi.scholarsportal.info/documentation/{*surveyID*}/{*filename*}
  - {*surveyID*} is the file name of the dataset (minus the file extension)

Click on '+' beside **Series Statement** to open section. Click on field name to edit.

c) Other Study Materials

Describes other related studies or publications for this data file. To enter a study or publication:

- **(a)** Highlight the section the survey is to go under
- **(b)** Click **'+'**
- **(c)** Enter title of study and any other relevant information



d) Datasets

The Dataset section shows all the datasets imported into the file.

Click on the '+' beside the dataset you want to edit.

i) File Description

- **(a) File Contents**
  Brief Information about what the data covers
- **(b) Data Check**
  Information on any checks performed on the data
- **(c) Version Statement**
  - **(i) Version Text**
    Version number (if there is no number, use the date)
  - **(ii) Version Date**
    Date of revision
  - **(iii) Version Responsibility**
    Who is produced the version (i.e Statistics Canada)
  - **(iv) Version Notes**
    Reason for revision (e.g. errors in processing re-weighting)

ii) Variables

Resequence and declare weights if not already done.

- **(a)** Check variables (most of this should have been done in SPSS)
  - **(i)** Variables and value labels are correct
  - **(ii)** Missing values are declared (if not, dataset will have to be corrected in SPSS and re-imported)
- **(b)** Question Text / Universe / Notes
  Click on **Documentation** tab in middle of screen and fill in fields. Cut and paste from documentation (user guide, codebook, questionnaire, etc.)

iii) Data Entry

This section shows the data. Do not change anything here.

e) Variable Groups

Variable groups cluster variables with similar characteristics.

i) Create Variable Groups

(a) Click on **Description** tab

(b) On left-hand side, click on **Variable Groups**, then on '**+**'

(c) Type group name in **Label:** field (right side of screen) and any additional information about the variable group

(d) Next variable → click on **Variable Groups**, then on '**+**'

- if you click on '**+**' before **Variable Groups**, you will create a sub-folder of the current group; to remove folder, click on '**–**'

ii) Add variables to groups

(a) Click on **Variables** tab

(b) Highlight the Variable Group that you want to add the variables to

(c) Click on '**+**' on far right side of screen

(d) **Select Variables** window opens

(e) Choose variables to add, click **OK**

- a green dot appears beside a variable when it has been added to a group

(f) To remove variable from group → highlight variable → click on '**–**'

(g) To move variables up/down → highlight variable → click on ↓ or ↑

(h) When done, check the list to make sure all the variables in the dataset have been put in a group. If they are, they will all have a green dot beside them

f) Other Materials

Provides links to all documentation that accompanies the data file for this study: codebooks, user guides, questionnaires, SPSS and SAS program files, etc.

i) Click on '**+**'

ii) Enter title of study and location on the <odesi> ftp server

- Location of ftp site:
  http://odesi.scholarsportal.info/documentation/{*surveyID*}/{*filename*}
  - {*surveyID*} is the file name of the dataset (minus the file extension)

g) External Resources

External Resources is used for aggregate data, allowing it's metadata to be described the same way as a data file.

See **External Resources Guide** for instructions.

6) **Validate (Tools → Validate …)**

a) **Validate metadata**

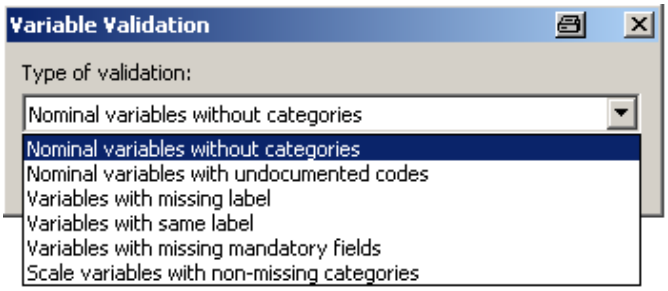
This option checks that all mandatory metadata fields are completed. If any empty fields are found they will be marked in red. Fields are marked mandatory in the Template Editor.

b) **Validate External Resources**

This options checks that all mandatory fields for 'External Resources' are completed. If any empty fields are found they will be marked in red. Fields are marked mandatory in the Template Editor.

c) **Validate Variables**

This option enables you to validate the variables within a dataset and to locate those that may have missing information, e.g. missing category labels.



The following options are available:

- **Nominal variables without categories**
  Checks for missing category labels within nominal variables.
- **Variables with missing label**
  Checks for any variables with missing labels.
- **Variables with same label**
  Use to quickly locate variables that have the same label.
- **Variables with missing mandatory fields**
  Checks for missing mandatory information.
- **Scale variables with non-missing categories**
  Locates scale variables that contain categories that are not defined as 'missing'. For example, some 'Age' variables record a respondent's exact age, but may include a category '99' labelled 'Not answered'.

To use:

i) Select variables to be validated by highlighting variable

ii) In the Variables section, **Tools → Validate Variables**

iii) Select the type of validation from the drop down list

iv) Click OK

A message will be displayed saying how many variables were found that matched the validation criteria, and these will be highlighted in the Variables window.

7) **Save file**

b) **File → Save As…**

c) Name file same as .sav file and save in appropriate directory → **Save**

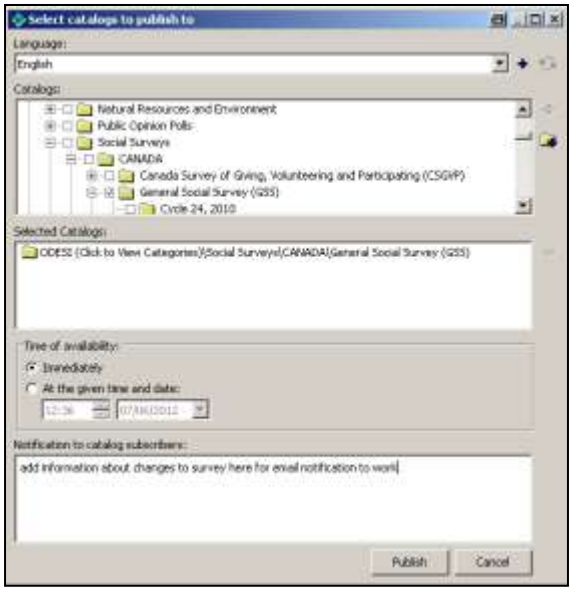
8) **Exporting Metadata (.xml file)**

Export the file to create the .xml file. Make sure that you have saved the file before exporting the metadata.

a) Click on **Documentation → Export → Export DDI…**

b) Save DDI file as .xml in proper directory

9) **Publishing Data and Metadata**



   a) **Publishing** → Server Name →**Study** → **Publish data and metadata…**
   b) *Select catalogs to publish to* dialog box → find folder that study belongs in →put checkmark in box beside folder name
      i) If a folder does not exist for a file, create a new one in **Manage Server…**
      ii) Check to make sure that no other folder is selected; Publisher will sometimes 'remember' previously selected folders
   c) If the file is not to be made public immediately:
      i) Under **Time of availability:** → check **At the given time and date:**
      ii) enter time and date
   d) **Email Notification**
      If substantial changes have been made to the file and notification needs to go out:
      i) in the **Notification to catalog subscribers:** box, list changes made to file
   e) Click on **Publish**
   f) Window opens **– Working, Please Wait…** (some files can take a while to publish)
   g) When file is published, **Publishing Report** will open → click on **Open in Web client…** to view and perform quality checks on final published product.

10) **Republish**
   Only use this for problems that can be fixed in Publisher; if problem is fixed in SPSS, you must re-import the file.
   a) **Publishing** →Server Name → **Study** → **Republish**
      - study will publish to folder(s) originally published to

      OR

      Publishing → **Republish on all servers**
      - this will publish the file to any of the servers it has already been published to
   b) Window opens **- Working please wait…** (some files can take a while to publish)
   c) When file is published, **Publishing Report** will open →click on **Close**

RE-IMPORTING THE DATA FILE INTO NESSTAR

If there is a problem with the data file that cannot be fixed in Publisher, you will have to fix it in SPSS and re-import the datafile.

Before you do this, ensure that you have exported the metadata as described above. Once you have finished fixing the data file, re-import the data following the steps below.

a) Import dataset
b) Resequence variables
c) Import .xml files
   - uncheck Variable Information – if you leave this checked, the .xml file will override the dataset and the changes made in SPSS will be gone.
d) Save .Nesstar file
e) Export Metadata (.xml file)
   - exporting the file creates the new .xml file.
f) Re-publish the file.