

# **Publishing Synthetic Data Files to <odesi> Using Nesstar Publisher 4.X**

March 2015

Jane Fry, Carleton University

Carys Carrington, Carleton University

**Note: Synthetic data files are not real data, that is, they are not the same as microdata or aggregate data files. Therefore, they require different mark-up procedures.**

### **SPSS**

- Open the dataset in SPSS.
- Eyeball the data to make sure that every variable has a label.
- Apart from the above step, nothing else has to be done in SPSS.

### **FTP**

- All synthetic files have a separate folder in FTP under the appropriate Survey folder.
  - e.g. ....*/documentation/CCHS\_Synthetic/2012/*
- Place the SPSS file on the FTP server in the appropriate folder.

### **Nesstar Publisher**

- Open Nesstar Publisher
- **DO NOT** “Import Study” into NESSTAR (that is, the dataset)
- Go to
  - *File*
  - “Add New Study”
  - Begin creating the metadata.
- Remember to save your file regularly!
- Follow the “How to” Guide for entering the metadata **EXCEPT** for the following tags:
  - **Title**
    - All titles, subtitles, alternative titles start with
      - ♦ “Synthetic: Title”
        - e.g. *Synthetic: Aboriginal Children’s Survey, 2006 [Canada]*
  - **Parallel Title**
    - Do **not** enter a parallel title.
  - **ID**
    - Follow the formula given in the “Best Practices Document (BPD)” ***with*** the addition:
      - ♦ at the beginning of the ID, add the word “***synthetic***”
        - Example 1: *synthetic-cchs-E-2012-mental-health*
        - Example 2: *synthetic-nlscy-E-1998-1999-c-3-primary-file*
  - **Bibliographic Citation**
    - Follow the formatting as given in the Best Practices Document (BPD) ***with*** the addition indicating that it is a synthetic file.
      - e.g. *Statistics Canada. 2009. Synthetic File: The National Longitudinal Survey of Children and Youth [Synthetic File]. Ottawa, Ontario: Statistics Canada. Special Surveys Division [producer and distributor].*


- **Date of Collection**
  - Do **not** enter the “*Date of collection*”.
- **Abstract**
  - The following paragraph has to be added to the Abstract.
  - It should **precede** the description of the survey.

“Please note: This is a Synthetic data file, also known as a Dummy File – it is **NOT** real data. This synthetic data file should **not** be used for purposes other than to develop and test computer programs that are to be submitted by remote access. Each record in the synthetic file matches the format and content parameters of the real Statistics Canada Master File with which it is associated, but the data themselves have been ‘*made up*’. They do **NOT** represent responses from real individuals and should **NOT** be used for actual analysis. These data are provided solely for the purpose of testing statistical package ‘code’ (e.g. SPSS syntax, SAS programs, etc.) in preparation for analysis using the associated Master File in a Research Data Centre, by Remote Job Submission, or by some other means of secure access. If statistical analysis ‘code’ works with the synthetic data, researchers can have some confidence that the same code will run successfully against the Master File data in the Research Data Centres.”

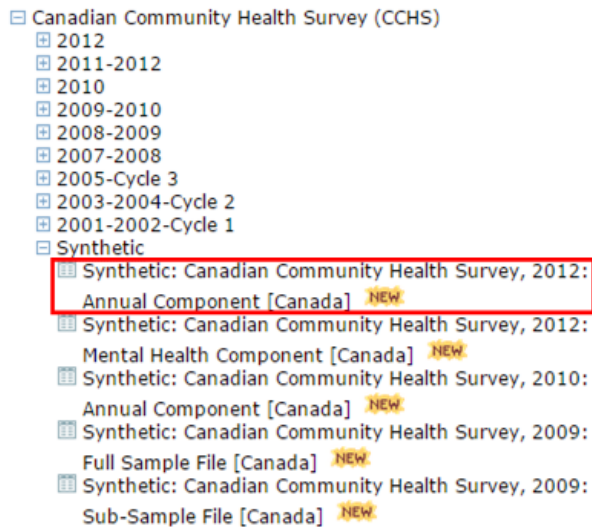
- **Kind of Data**
  - Synthetic Data
- **Time Method**
  - Do **not** enter the “*Time Method*”
- **Data Collector**
  - Do **not** enter the “*Data Collector*”
- **Sampling Procedure**
  - Do **not** enter the “*Sampling Procedure*”
- **Mode of Data Collection**
  - Do **not** enter the “*Mode of Data Collection*”
- **Weighting**
  - Do **not** enter the “*Weighting*”
- **Locations of Data Collections**
  - To direct users to the location of the data
    - ◆ add the link “*Access data here*”
    - ◆ the URL for the link will be to data that is in the folder on the FTP server
- **Variable Groups**
  - Variable groups do not have to be added to the synthetic files (yay!)

- **Publishing the survey in ODESI**

- Create a separate Folder named “*Synthetic*” under the appropriate survey name on both servers.
  - The “*Synthetic*” folder should be the last folder listed under the appropriate survey.

e.g. 

- Put the completed synthetic file within the “*Synthetic*” folder.
  - e.g. *Health/Canada/CCHS/Synthetic/Synthetic: Canadian Community Health Survey 2012: Annual Component*



- **Note:** Ensure that only the appropriate folder is checked off prior to publishing the file on both servers.