<u>Standard Operating Procedure at the National Treasury Secure Data Facility</u>

1. **In-file requests**: Researchers need to send an email to sa.datateam@wider.unu.edu with the files they would like to place under their 'in_files' folder on the server. Most in-file requests consist of do files.

2. **Out-file requests:** Researchers will need to request output. The process is as follows:

   a. Researchers should send an email to sa.datateam@wider.unu.edu requesting their output.

   b. In the email, researchers need to state the exact file directory of their output request. We recommended that researchers place their output in a folder labelled by the current date in their 'out_files' under a researcher's workbench. For example, **Z:\Workbenches\'ACCESS'\name_surname\out_files\yyyymmdd**

   c. The email request should also include a brief explanation of the output (for example, regression results or summary statistics by industry code).

   d. It is the responsibility of the researcher that the requested output is in line with the data lab's data policy:

      i. Output that identifies individuals or firms may not be sent to researchers. Cells that contain less than 10 observations will not sent and should not be requested. This would contravene the NDA agreement signed to gain access to the data (anonymity of firms and individuals is of the utmost importance).

      ii. The most granular information that may be requested from the data lab can be spatially aggregated on the provincial level and by the 1-digit industry code, subject that cells contain more than 10 observations. (For datasets that currently do not have 1-digit industry codes, these can be merged in by researchers using the industry concordance tables found in this file path "D:\Researchers\Master Data\Industries\sic5_concordance"

      iii. Please do no export Stata graphs (.gph format) as these graphs include the underlying data which is a security risk and are very large to transfer.

iv. Researchers should refrain from encrypting the data with the hope of exporting the encrypted datasets. The data team will not export encrypted datasets from the data lab.

e. We aim to send the output for out-file requests within 24 hours. Any out-file requests received over weekends will be attended to in the following week.

3. **Example of output that can potentially reidentify the data and for that reason will not be sent to researchers:**

For example: Investigation of SIC code 01 by taxable income:

Tax bracket	Number of observations	Industry	Does this violate confidentiality?
0-5000	3	Agriculture	Yes
5001-10000	9	Agriculture	Yes
10001-15000	25	Agriculture	No
15001-20000	80	Agriculture	No

4. Example of aggregated datasets/tables that will not be sent to researchers

Province	3-digit Industry	sales	cost of sales	capital
western cape	MINING OF GOLD AND URANIUM ORE	5777	7704	4033
western cape	MINING OF IRON ORE	4832	9889	5591
western cape	MINING OF NON-FERROUS METAL ORES, EXCEPT GOLD AND URANIUM	5239	3378	9891
western cape	STONE QUARRYING, CLAY AND SANDPITS	8103	8002	5352
western cape	MINING OF DIAMONDS (INCLUDING ALLUVIAL DIAMONDS)	6024	4610	8329
western cape	MINING AND QUARRYING N.E.C.	3221	5739	7149
western cape	SERVICE ACTIVITIES INCIDENTAL TO MINING OF MINERALS	4195	5109	8776