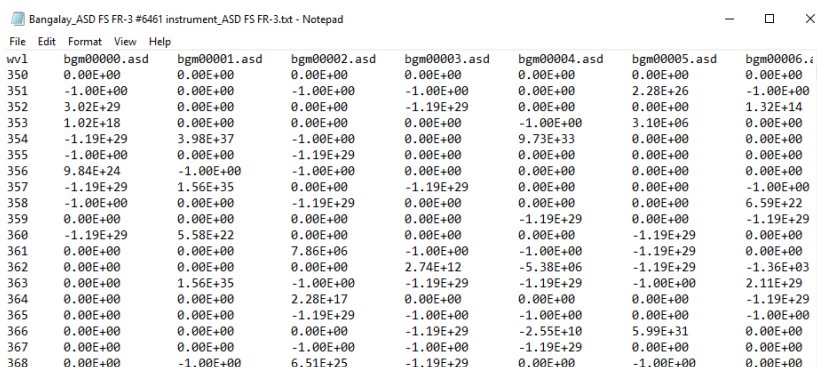**National Spectral Database Write-access Guide**

Last updated: 10/06/2021

Use this guide in accordance with the NSD General Guide, Metadata Standard and Terms of Service to add data to the NSD. If you are unfamiliar with these documents, please review them prior to attempting to add data.

To be able to add data, please indicate your institute or company name when creating a NSD account via the Specchio client, and provide a brief summary of your dataset along with a suitable campaign name via email to the NSD manager (NSDB_manager@ga.gov.au). The NSD manager will create a root folder for you on the database with the campaign name as folder name, and from there you may populate the folder from the client: Data input > Load campaign data, select your campaign name and target data folder.

Uploading data

Specchio supports ingestion of a range of raw data types such as ASD binary files, and this is done by selecting "data input". CSV should be converted to a text file for the main data body (raw wavelength information) and separate .xls file for the metadata / header info. Raw data in a text format should look like:

This text file was generated by selecting a subset of data from a .csv and saving as a tab-delimited .txt file. It is important to have "wvl" in the top left hand position as above, as this is required by Specchio so that it recognises the text file. Also note file names must be in the first row, with data for each file below in the relevant column. Spectral wavelength must be in the first column, under the "wvl" heading, with measurements in subsequent columns under their file names (long data format). Alternatively, .xls file format are also supported for raw data ingestion.

File sub-folder structure should be considered when ingesting into the NSD. Classifying by study type then location (if applicable) and then the spectral data itself, with an institution as the overall base folder (also where applicable) is the preferred folder sub-structure. If you represent an institution, this should be the name for your base / root folder, and an appropriate base name can be decided upon with the NSD manager. If your study is a one-off, please select an appropriate, descriptive folder name such that it is identifiable to you, and sensible enough to be distinguished by someone without prior knowledge of your study. Where an over-arching institute isn't available or applicable, a project name or description can be used in place for the base folder name.

Note: The name you supply as your campaign name for uploading a new campaign is used as the root folder name in Specchio. The folder you specify to upload from will be inserted at the next sub-folder level within this campaign folder, so please take care to create a sensible sub-folder structure. You can add new folders to your campaign at a later time, either empty or with data in them by selecting Data Input > Load campaign data from the Specchio client. Select your campaign to add data to, then specify the folder location (either use the existing path or select a new one) and select "Load". You can also move data at a later time, by selecting Data Maintenance > Move data, navigating to your campaign and selecting the folder you wish to move. You can only move data within your campaign, not between existing campaigns. You can "change" folder names by uploading an empty folder with the desired name and shifting data into it from the client (Data Maintenance > Move data).

Please also ensure that appropriate metadata fields are populated for your dataset, such that others may find your spectra based on metadata parameters including location, time and institution or study name.

Geoscience Australia's ARD cal/val field data folder sub-structure could look like:

-Geoscience Australia (campaign name)

    -Field data (the folder that you select from the file browser within Specchio)

        -Site (ie Lake George)

            -Date (of field survey)

                -Files (ie .asd1, .asd2 …)
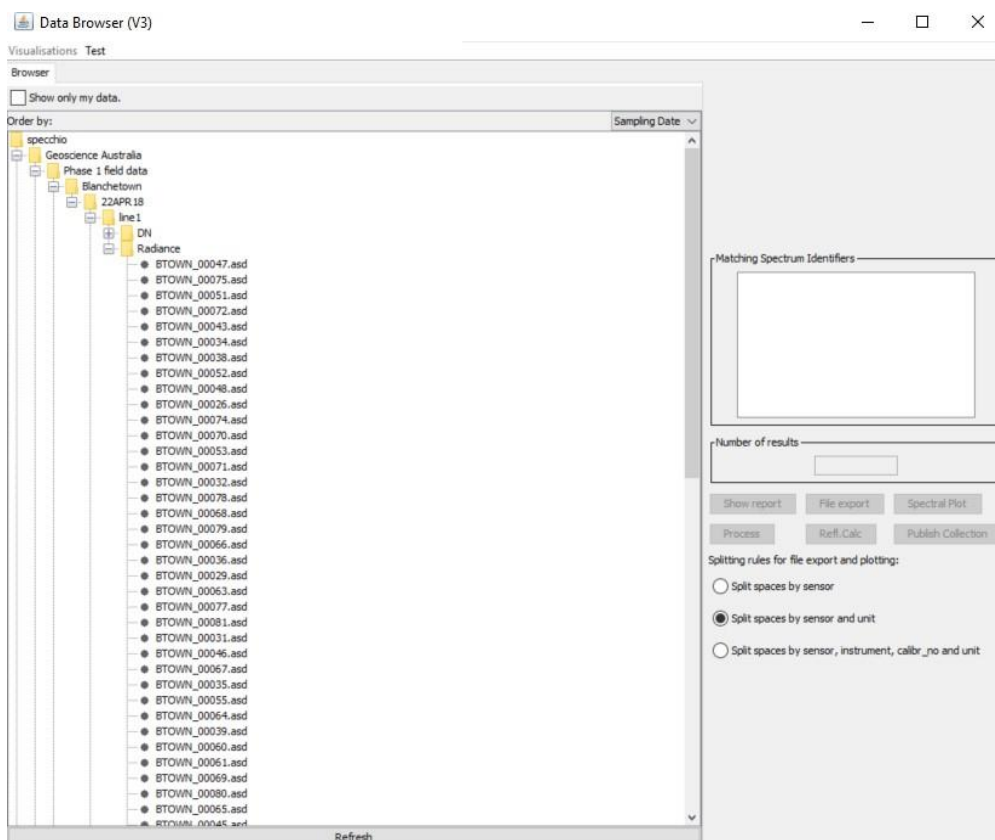
And if we had CSIRO substrate data it could look like:

-CSIRO

    -Substrate data

        -Site

            -Files

Further information and an extensive guide to supported file types, import processes and supported metadata parameters are available in the Specchio documentation: https://specchio.ch/guides/

Uploading metadata

Populating relevant metadata fields minimises reliance on the file structure in the national spectral database. An internal query-builder can be utilised to pull out desired spectra, given a metadata search.

Specchio supports Excel .XLS or .TXT metadata imports, with columns matching the file names then metadata fields you wish to upload (wide data format). An Excel workbook can be saved as an .XLS file (make sure the correct tab is selected / in the 1st tab position in Excel when saving). An example setup of a metadata .XLS is shown below:

File Name	Temp	Lat	Long	Elevation	etc
001.asd	22.8	-35.2	-141.1	45	
002.asd	23.1	-35.2	-141.1	45	
003.asd	23	-35.2	-141.1	45	
etc					

- To import metadata for a given study / campaign, within Specchio select "data input", "get metadata from XLS". This will open a new window, select the folder that you wish to apply metadata to.
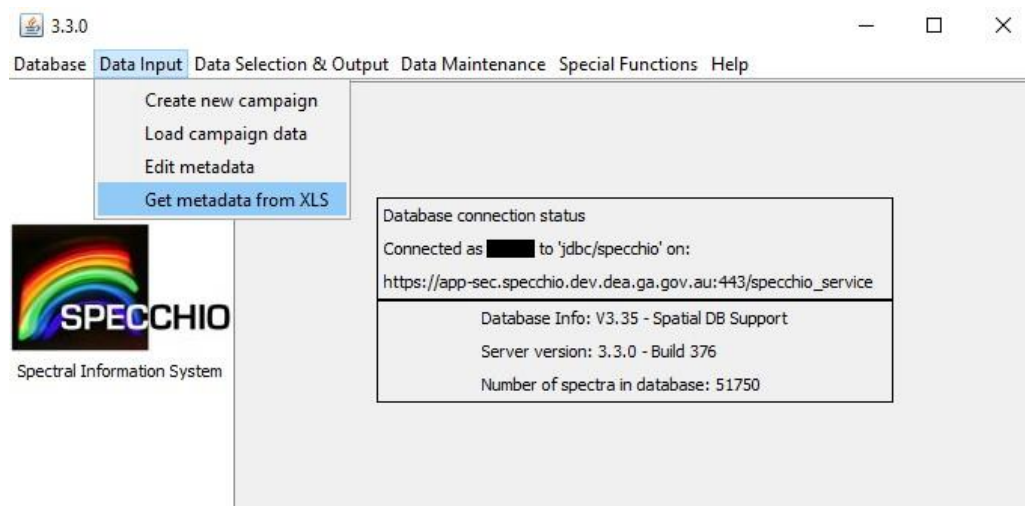

- To specify the .xls metadata, select "file" at the top left hand side of the window, above the folder panel. This will open up a matching element control window on the RHS, and this is where you can specify where the metadata fields are populated within Specchio.
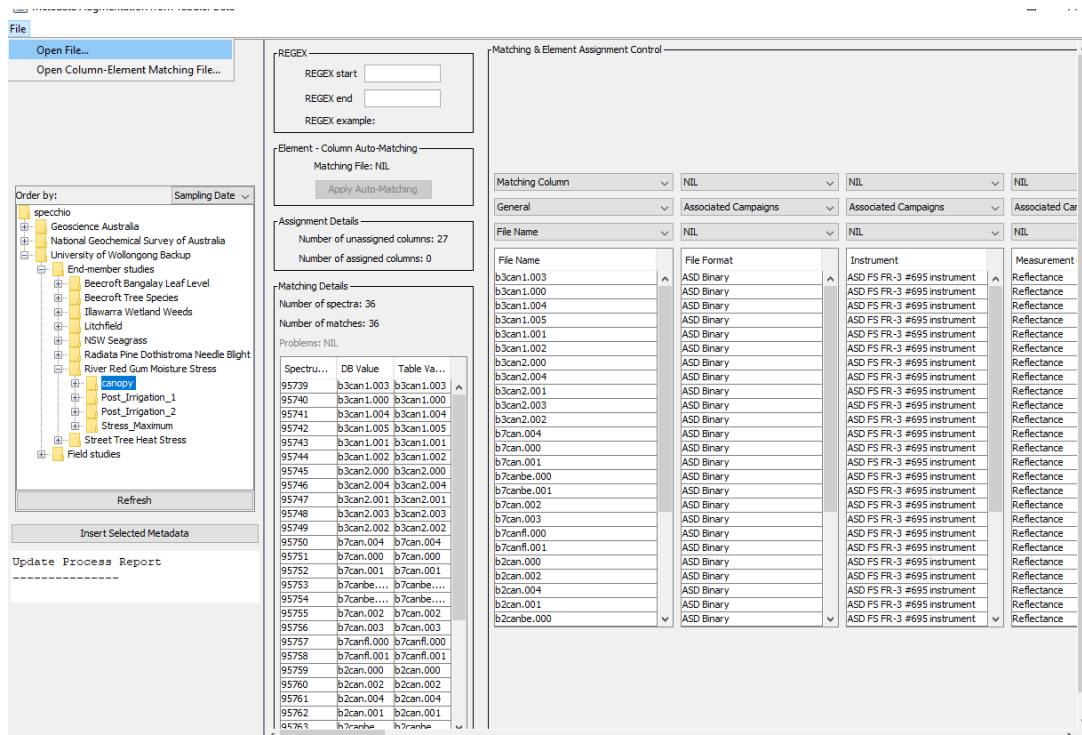
- The first thing to do is specify a "matching column" which is a unique identifier (the file name). Scroll across to the column which includes the file names, and in the three drop down lists above specify "matching column", the metadata category "general" and the metadata parameter "file name". This will tell Specchio to associate the rest of the metadata based on file name.

- The remaining metadata columns can be specified in a similar way, but leave the 1st drop-down list as "NIL" as we have already specified a matching column. For example, a "calibration ID" column will need to be added to the related metadata field in Specchio which is "instrument" in the 2nd drop-down list, and the specific metadata parameter is "calibration number". A comprehensive list of metadata categories available in Specchio is available in the Specchio user manual.

- If importing metadata for multiple studies / folders you can save time by using a column matching XLS file. This must have the first row as shown below, with your column names in the first column and the Specchio metadata parameter in the second column (metadata parameter names must be an identical match).

	XLS Column Name	Specchio Name
1	XLS Column Name	Specchio Name
2	File Name	File Name
3	File Format	File Version
4	File Comments	File Comments
5	Instrument	Extended Instrument Name
6	Measurement unit	Raw Data Format
7	Sensor	Sensor
8	ANDS Collection Description	Target Description
9	ANDS Collection Name	Data Usage Policy
10	ANDS Collection Key	Data Usage Policy
11	Acquisition Time	Acquisition Time
12	Automatic Dark Current Correction	Automatic Dark Current Correction
13	Campaign Name	Campaign Name
14	Citation	Citation
15	Common	Common
16	FOR Code	FOR Code
17	FOV	FOV
18	File Version	File Version
19	Gain_SWIR1	Gain_SWIR1
20	Gain_SWIR2	Gain_SWIR2
21	Integration Time	Integration Time
22	Keyword	Keyword

- A list of current metadata fields is most easily accessed from the Specchio client. Select Help > List available Metadata Elements. Copy desired metadata fields into the "Specchio Name" column of your matching-element XLS file, and add the column names you use in your metadata file in the "XLS Column Name" column, next to the appropriate Specchio field.

- Select file, open column-element matching file and select the file you generated. Select "apply auto matching" and if done correctly below this button your columns should be assigned.
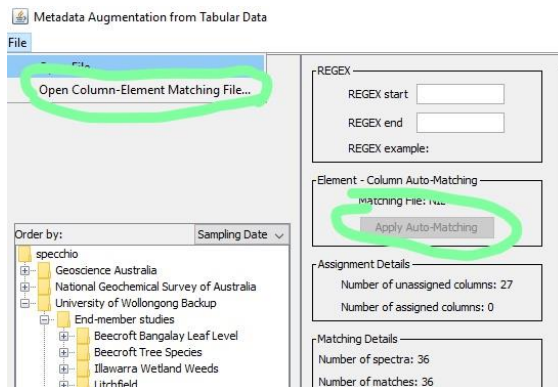
- Once you specify a matching column (file name) the matching details below should populate, and you can select "insert selected metadata" on the left of the window, a dialogue box will appear with the status of the upload. Using a column matching file when uploading metadata is the quickest way to select the metadata fields you wish to populate for your dataset.

Note: You can also manually edit metadata from Data Input > Edit metadata. This brings up a browser, with the Specchio folder structure on the left and metadata fields on the right. Generally, you will only be interested in editing Spectrum level metadata, so select this tab at the top. Also select the files(s) or folder you wish to alter the metadata for (changes will be applied at the level you select, ie for a single file, sub-folder of for an entire campaign). Right-click in the appropriate metadata category to see a list of available new metadata fields, or click on an existing entry to alter it. For each change, select "Update" at the bottom-left of the window to apply the change. If you do not do this after every update, changes may not be applied to the database. The image below shows selecting a new field, "Study Description" from category "Campaign Details" to add to the file "4" from a test data folder. Once you select the field, you will be able to alter it then apply the change.
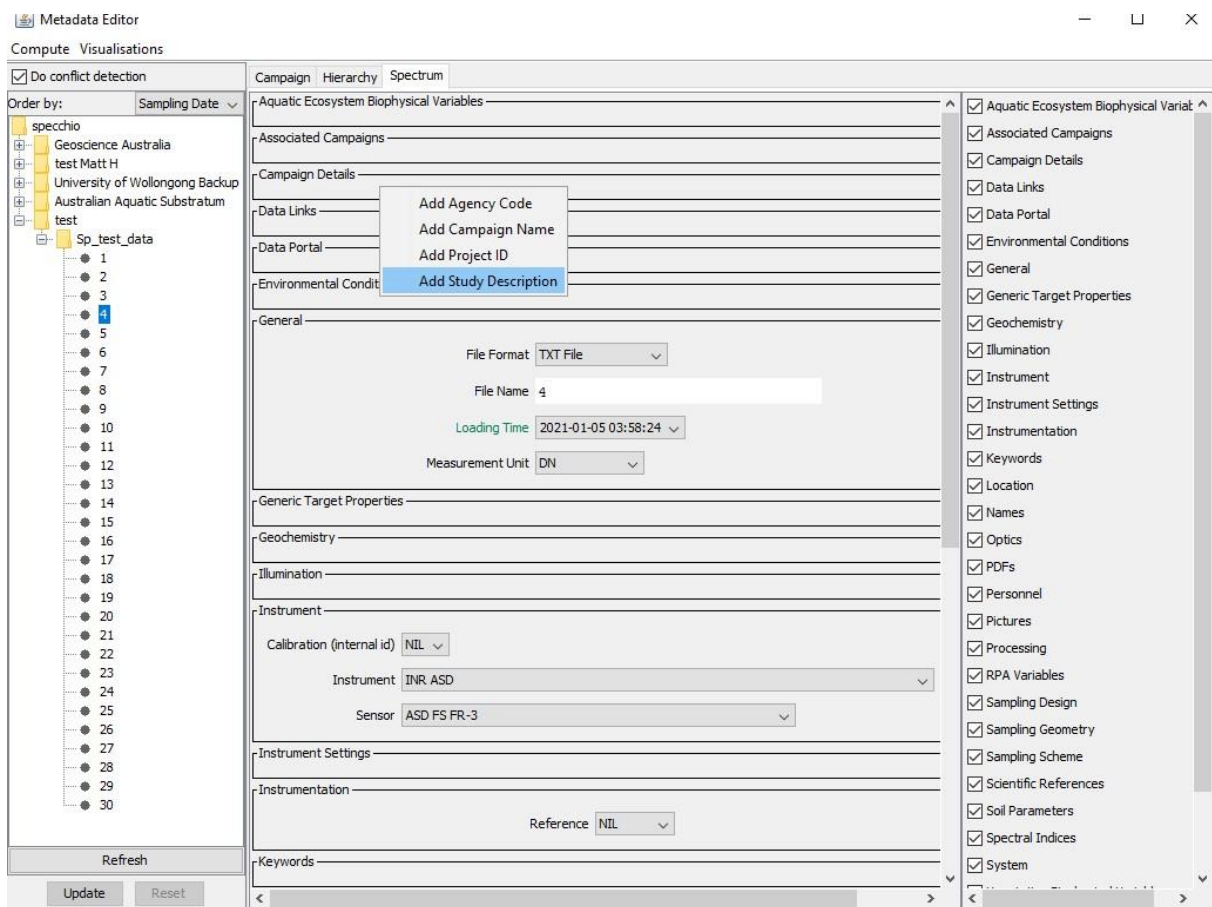
Further information related to campaign import and file types supported by Specchio are described from pg. 66 (Section 9, Data Input) and pg. 29 (Section 7, File Formats) of the Specchio user manual which is available from the Specchio client website:

https://github.com/SPECCHIODB/Guides/raw/master/SPECCHIO_UserGuide.pdf

Information related to metadata import can be found from pg. 81 of the Specchio guide including the .XLS column-matching process.

Python, R & Matlab access *advanced users only*

Specchio NSDB access can be achieved outside the client, via Python, R or Matlab scripts. You can also build a Matlab engine into Python, such that Python runs Matlab scripts and can be built into existing workflows. You will need server credentials and to prepare a script in your preferred language to connect to the database. You may then query the database to pull out desired spectra, and can plot and download subsets of spectra and their metadata as needed. A good way to retrieve code that queries the database is via the client, and by right-clicking within the "matching spectrum identifiers" box after you have run your query.

Alternatively, it looks like this in R:

```
query$setQueryType(query$SELECT_QUERY)
query$addColumn("spectrum_id") # <- what you are querying for#
cond <- .jnew("ch/specchio/queries/QueryConditionObject", "spectrum", "spectrum_
```

```
id");
id_array <- c(72,73,74,75,76,77,78,79,80,81) # <- matching spectrum identifiers#
ids_list <- .jnew("java.util.ArrayList")
for (i in 1:length(id_array)) { ids_list$add(.jnew("java.lang.Integer",
as.integer(id_array[i]))) }
cond$setValue(ids_list);
cond$setOperator("in");
query$add_condition(cond);

ids <- specchio_client$getSpectrumIdsMatchingQuery(query);
```

This code returns the spectrum id's that match the spectrum identifiers supplied. Further documentation is available from https://specchio.ch/guides/ which covers Matlab and R access, and how to plot and download data.

Python access can be achieved via the JPype package: https://jpype.readthedocs.io/en/latest/#

You will need to specify the correct server credentials and login info to gain access.

A detailed guide for Python access to the NSDB is available at:

https://github.com/SPECCHIODB/Guides/blob/master/Programming%20Course/Python_accessing_SPECCHIO.pdf