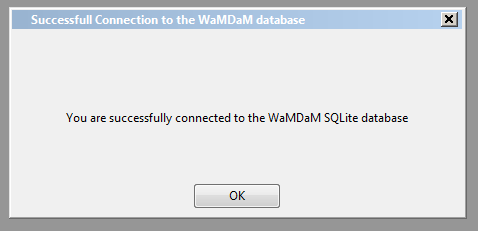
**Data Validation, integrity checks, and transaction management**

**Adel Abdallah- Jan 4, 2017**

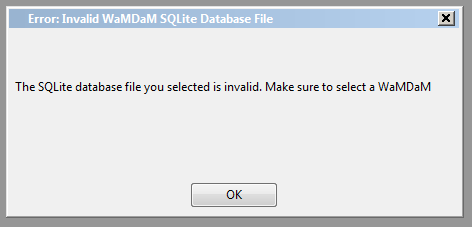
The purpose of validations and interactive messages in the WaMDaM Wizard is to provide a user’s-friendly experience that informs the users of their successful or erroneous actions while using the Wizard. The validations will be implemented at four different steps: i) connect to the database, ii) read the workbook/Hydra/Data-Preparation-sheets, iii) parse the tables or input data and map it to the WaM-DaM database, and iv) load data to the WaMDaM SQLite database. The process from connecting to a database up to loading data into WaMDaM and any possible errors will be documented through WaMDaM.log text file.

1. Connect to the database

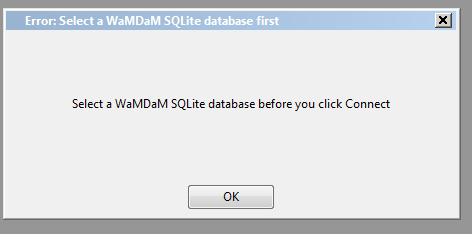
Check if the selected database file is (a) a SQLite3 file, and (b) it conforms to the WaMDaM database structure and version. For example, a WaMDaM database must contain all the following tables: Datasets, MasterNetworks, ObjectTypes, Mapping, Methods, and Parameters. The database can be blank and it can be populated with data. If this check is valid, show the following message “You are successfully connected to the WaMDaM SQLite database”.



If not then show the following message “The SQLite database file you selected is invalid. Make sure to select a WaMDaM database file. You can download a blank SQLite WaMDaM file from GitHub @ <https://github.com/amabdallah/WaM-DaM>”.



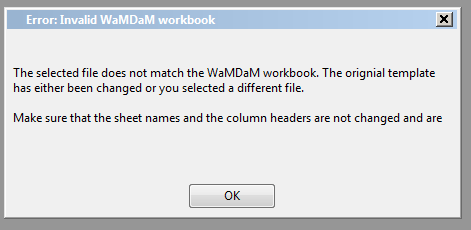
If the user clicks on “connect” to the database before selecting a SQLite file, an error message will show “You need to select a WaM-DaM SQLite file before you click connect”. Perhaps, as an alternative to this issue and similar validation issues later in the document, the “connect” or “Load” buttons should be dimmed (like turned off) until the user selects a database or a workbook, or whatever file.



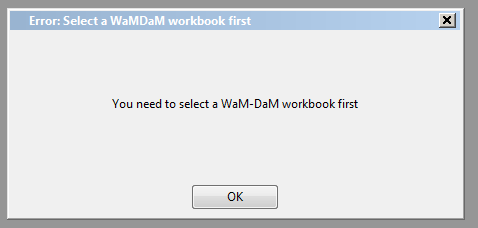
1. Read the workbook, Hydra, or any other input spreadsheet

Check if the selected workbook has the following items: i) a list of the spreadsheets that the Wizard will read. This list will not change and is provided in the template, ii) the “homepage” sheet in cell A3 has the value: WaMDaM Wizard version V1.0. [I might change this cell location or content later], iii) the location and headings of each spreadsheet/table read by the Wizard.

If any of the three above do not match with the default list, then return an error message: “The selected workbook file does not match the WaMDaM workbook template. The original template has either been changed or you selected a different file” You can download a blank WaMDaM workbook file from GitHub @ <https://github.com/amabdallah/WaM-DaM>”.



If the user clicks on “Advanced”, “Load data” buttons or unchecks boxes in the form “Import WaMDaM spreadsheet data” before selecting a valid workbook file, an error message will show “You need to select a WaM-DaM workbook first”

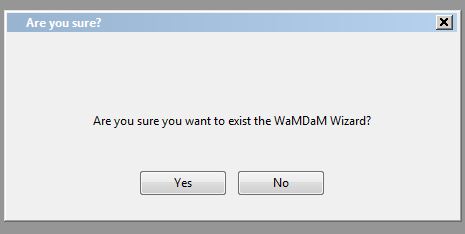


Advanced validation and maybe leave it for later

Check the numeric order of the “check boxes” in the Import Excel Basic form. If the database empty (no any metadata entries in the database), then the Data Structure check box should not be allowed to be checked if the metadata check box is not checked. In other words, the user has to check the metadata box first. This rule applies to the “Advanced” option in this form. Users cannot import data about “sources” until there is data already in the database about “People”. It gets tricky when there is already loaded data about “people” in the database but the user clicks to only import additional “sources” data. This will be an error of handling the violation of dependency or foreign keys when the user tries to load the data as I cover in the next section.

For the Hydra/shapefile/seasonal/time series import or data services options, the same last error message will show if the user tries to click load data before they select a file. Another validation check will apply to each source to check if the input file is valid given its format (excel) with appropriate headings, or xml file with the specific heading encoding.

When the user clicks to close/exist the WaMDaM Wizard, a message will ask them if they are sure if they want to exit. The two options are Yes or No. No will “destroy” this warning dialog box and “yes” will “destroy” the whole app.

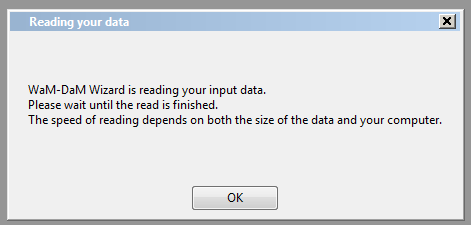


Advanced validation and maybe leave it for later

There should be an option to allow the user to save the connection to the database. If they click again to launch the WaMDaM wizard application installer, the Wizard skips the first step to connect to a database and automatically uses the previous database file. This save the connection should be added to the Connect to Database form. At the same time, the user should be allowed to change the existing database connection to a different database. This will be under File>connect/change database connection.

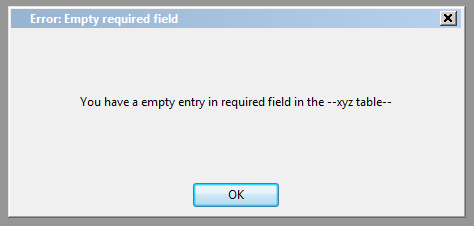
What if a user deletes of edits a table design in the SQLite database file later after connecting to it and loading data, there should be a check to see if the database still in the same design. Maybe there should be a quick validation once the user connects or revisits the database about all the available tables in the database. If one is missing, the Wizard should say that and refuse to load data.

A second level of check, is when the Wizard tries to load data to one single table at a time, the Wizard should check if the table in the database matches the original one, if something changed, similar message to above should up. I’m not sure how heavy the overhead would be to do these checks. Maybe will implement them later, but will keep them in mind.



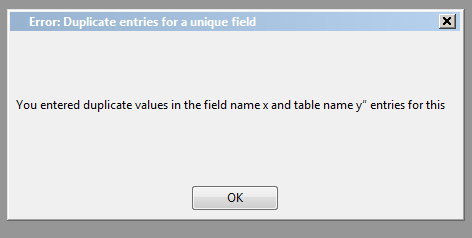
1. Parse the tables and map them to WaM-DaM

Upon reading data in each table, check if any of the fields that is marked as “required” is empty where there are data values in other fields. Also, check if the user skipped a row in excel and left it as blank. If yes, then the Wizard will show an error message: “You have a black entry in required field in the --Organizations table--”. The message must be specific to call which table has this problem. It night be easier to program the following dialog box to be dynamic to take the table name that has an error and display it as part of the message than using this static box from the wxforms.

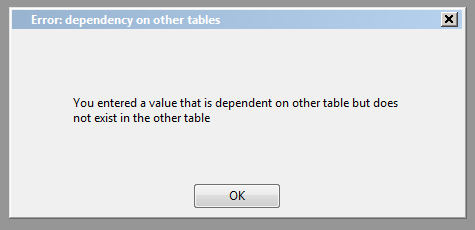


Check if the value entry in any table for a field violates the field’s physical data type in the database. Not sure when to do that. For example, if the user exceeds the length of field, the database will return an error while loading and messes things up. We need to be probative and prevent loading a bad data value by returning an error early on.

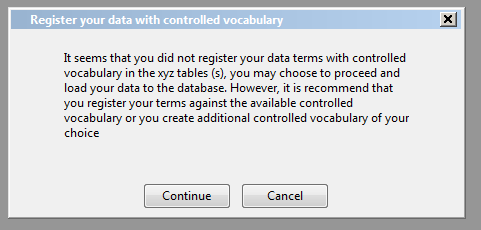
Check if the values are unique for the fields that are required and unique. If the user enters duplicate entries of a data structure acronym as an example, then a message will show up: “You entered duplicate values in the field name x and table name y” entries for this field and table should be unique. Delete the duplicated value”.



All foreign keys in mandatory related tables in Excel are required fields through look up tables. However, if a user for some reason manages to overwrite the look up value in excel with a different value that does not exist in the other depend table, then error message should tell them about violating a foreign key constraint like this “You entered a value that is dependent on other table but does not exist in the other table” the message should be customized to tell the user the names of both tables. This check will insure that users will populate tables in order. If they miss a look up value from another table that is empty. The error message will make them return and populate the missing and required field.

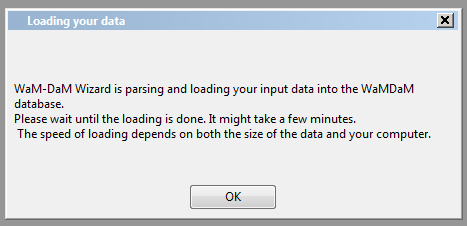


For controlled vocabulary optional fields, if the user leaves them empty in any table, that’s technical fine, but the Wizard will show a message to tell the user “It seems that you did not register your data terms with controlled vocabulary in the xyz tables (s), you may choose to proceed and load your data to the database. However, it is recommend that you register your terms against the available controlled vocabulary or you create additional controlled vocabulary of your choice” the message window will have two buttons: “Proceed and load data”, cancel, return to spreadsheet”

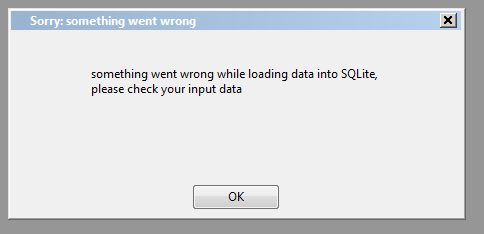


1. Load data to the database

When the Wizard loads data into the database, it should display this message



If for some reason, the database rejects the data while loading in the middle, ideally, we want to roll back then display this message: “something went wrong while loading data into SQLite, check your input data”. Ideally we want to avoid this error from happening because the user will have no clue why it happened. But just in case, we need a last resort error message for unforeseen bugs in the entered data.



If the spreadsheet passes these validations, then the Wizard loads the data values to SQLite. If the data loads successfully to the database, the Wizard should show this success message: “You successfully loaded the selected data into the database. You can view the data by using: SQLite Manager: Add-ons for Firefox web browser”.

