

Help with XML validation kits

This help document contains the following sections:

[Downloading the kit](#)

[Running the kit](#)

[Validation kit updates](#)

[Validation results](#)

[Errors tab](#)

[Schema errors](#)

[Warnings tab](#)

[Summary tab](#)

[Setting switches](#)

[Saving the error file](#)

[Configuration](#)

[Rules, results and switches folders](#)

[Max number of errors](#)

[Size of batch](#)

[Max viewable file](#)

[Use proxy](#)

Downloading the kit

This software runs the validation kits for the Student (Cyy051), Aggregate Overseas (Cyy052) and ITT In-Year (Cyy053) data collections.

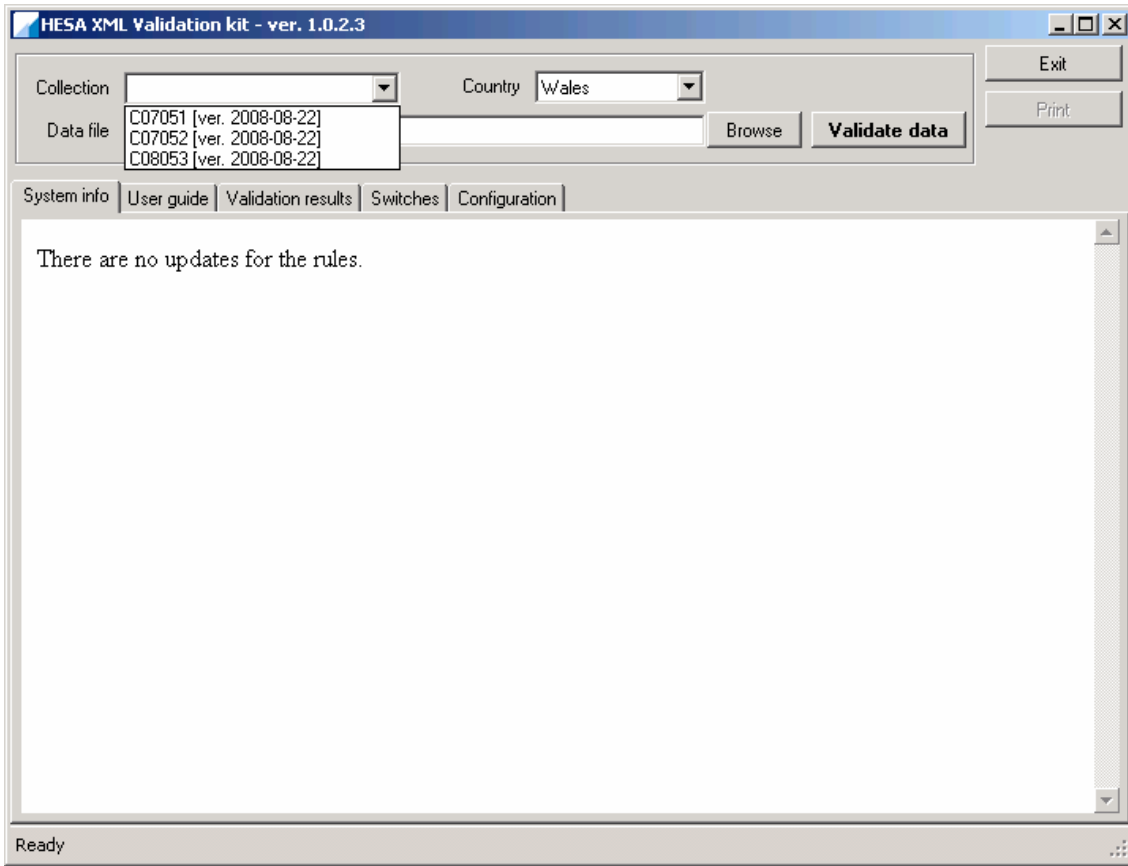
To download the kit, select the MIS installation file from:

http://www.hesa.ac.uk/index.php?option=com_content&task=view&id=1159&Itemid=268. This will open the setup wizard. Follow the step-by-step instructions of this wizard.

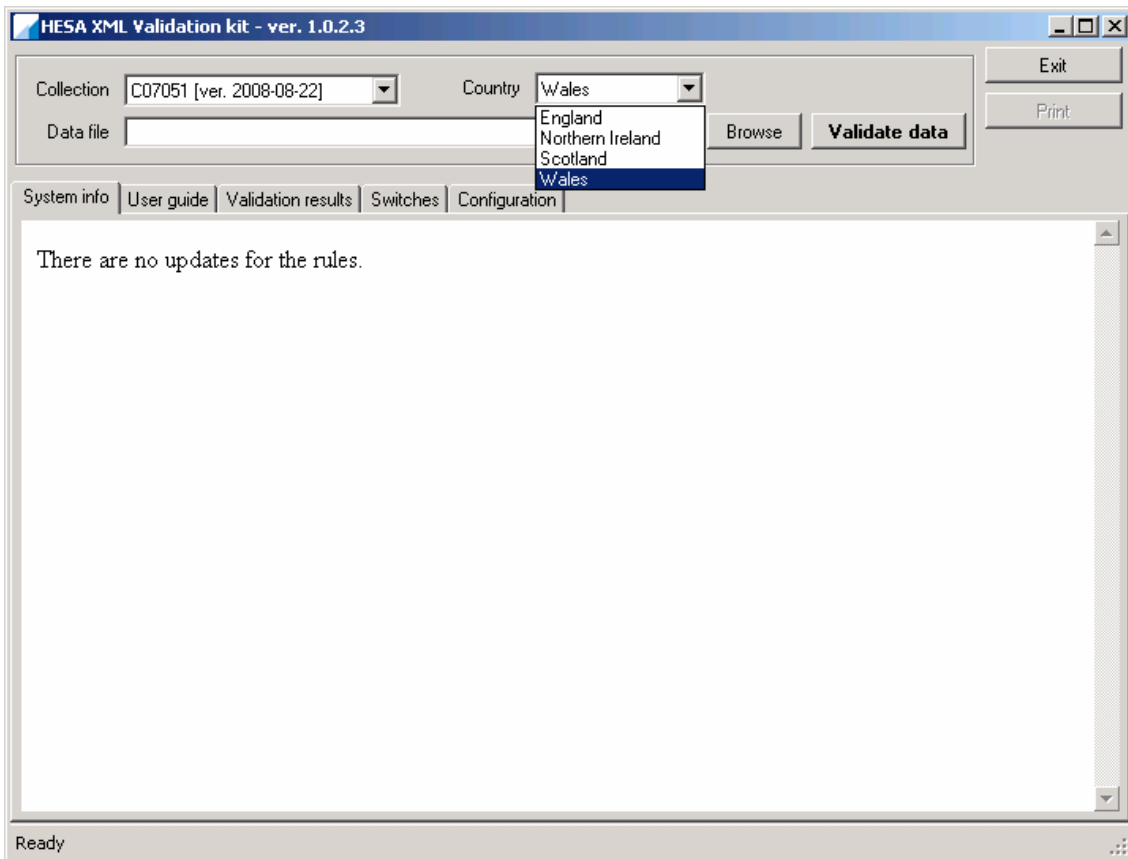
Running the kit

Double-click on the validation kit icon located on your desktop, or alternatively select and open the file from its saved location on your PC.

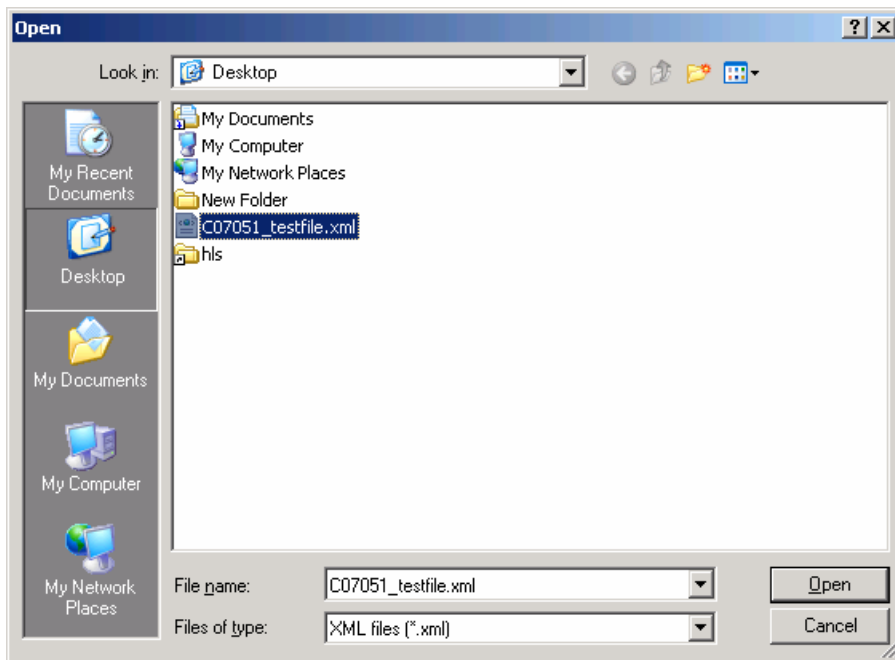
From the 'Collection' drop-down choose the collection you wish to validate: Student Record (Cyy051), Aggregate Overseas Record (Cyy052) or ITT In-year Record(Cyy053).



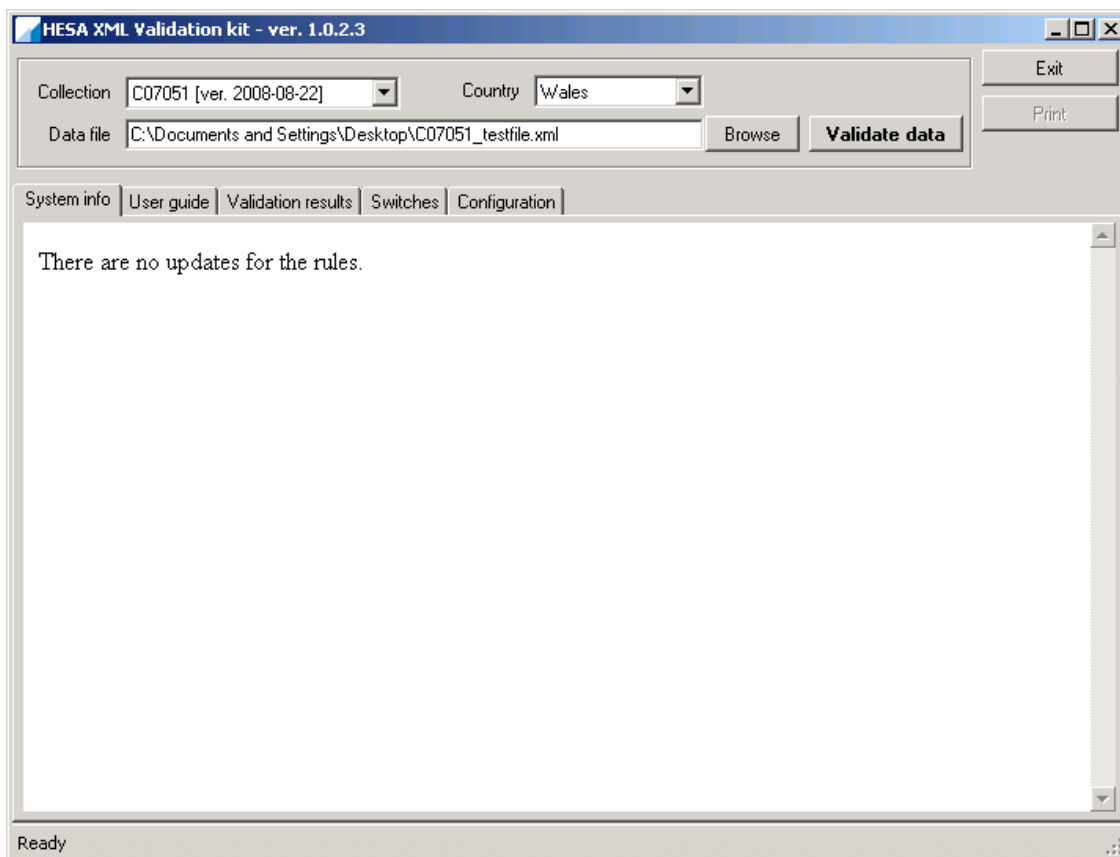
From the 'Country' drop-down select the country in which your institution is located. You will only need to select the country of your institution the first time you use the kit. Subsequently the kit will remember your selection.



Select the 'Browse' button, locate the file you wish to validate and click 'Open'



To validate the file click the 'Validate data' button.



Please note that the time taken by the kit to validate a data file will depend on both the size of the file, and the [batch size](#) you have specified in the Configuration Tab.

Validating data will activate three new tabs in which the validation results are displayed: Errors, Warnings and Summary.

Validation kit updates

When you start the validation kit it will automatically check the HESA server for updates, downloading these automatically where available. If new validation rules are available the kit will inform you that these updates have been applied automatically. If a new version of the kit is available then instructions for installation of the update will be provided. Details of downloaded updates will be listed on the 'Systems info' tab.

Validation results

The validation kit runs two stages of checks against your xml data file; schema rule checks and business rule checks. The validation kit will only progress onto the business rule checks once the file is clear of schema rule checks. Note that switches can only be applied to business rules.

Errors tab

The Errors tab details individual records failing errors which are required to be resolved before the file will pass INSERT-stage validation.

Collection: C07051 [ver. 2008-08-22] Country: Wales

Data file: C:\Documents and Settings\nicola_clark\Desktop\C07051_testfile.xml

System info | User guide | **Validation results** | Switches | Configuration

Errors | Warnings | Summary

Checking business rules

Validate each student against the business rules.
Records processed: 1

[number of errors: 15]

Errors

Check the box to switch off the rule for the given record

Off	Rule	Description	Location	Data
1	<input type="checkbox"/> Course.COLLOG.2	Course.COLLOG must not exist for institutions not in England	COURSEID = ENGLISH01	
2	<input type="checkbox"/> Course.CLSDCRS.2	Course.CLSDCRS must not exist for institutions not in England	COURSEID = ENGLISH01	
3	<input type="checkbox"/> Course.MSFUND.11	For institutions in Wales Course.MSFUND must not be coded 01, 03, 04, 05 or 07	COURSEID = ENGLISH01	MSFUND = 01
4	<input type="checkbox"/> Module.MODLANG.1	Module.MODLANG must exist for institutions in Wales, Scotland or Northern Ireland	MODID = E01	MODLANG =
5	<input type="checkbox"/> Module.LANGPCNT.1	Module.LANGPCNT must exist for institutions in Scotland, Wales or Northern Ireland	MODID = E01	LANGPCNT =

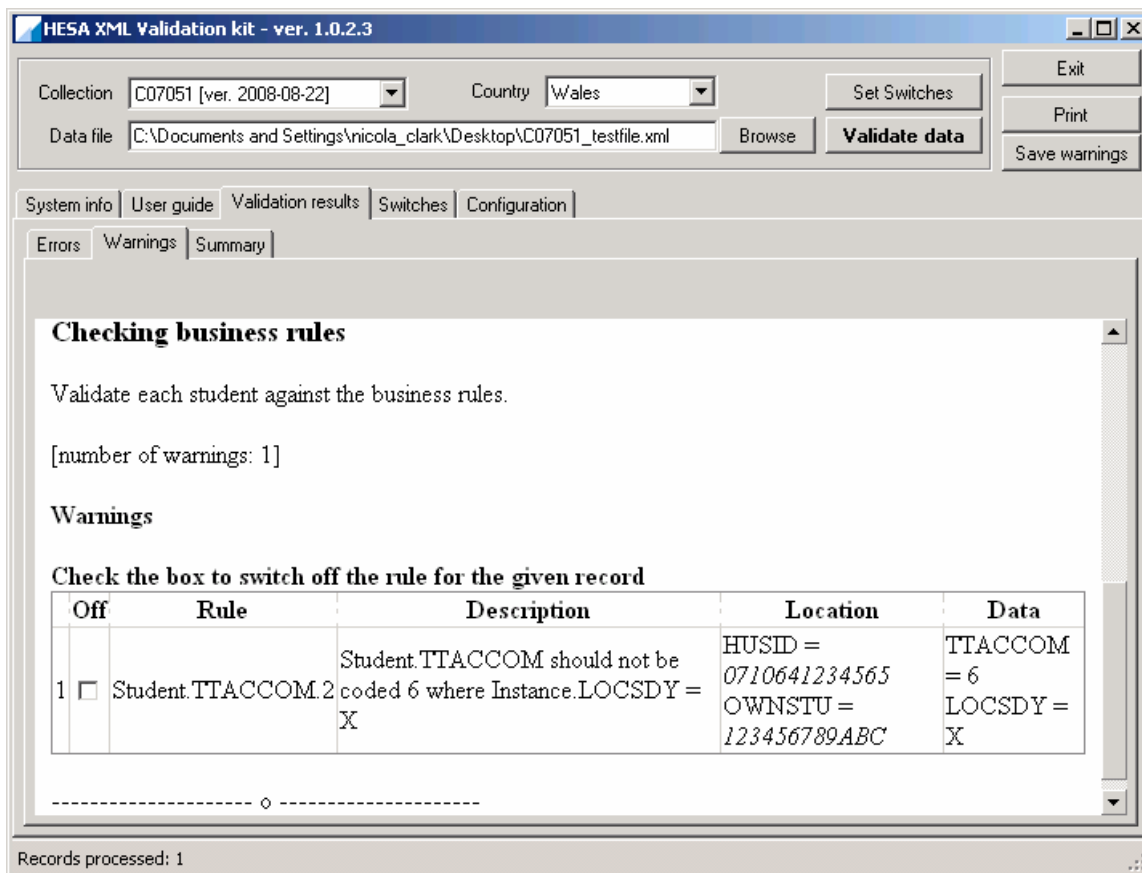
Records processed: 1

Schema errors

The validation kit will generate schema errors where the submitted xml file does not confirm to the schema (.xsd) of the record.

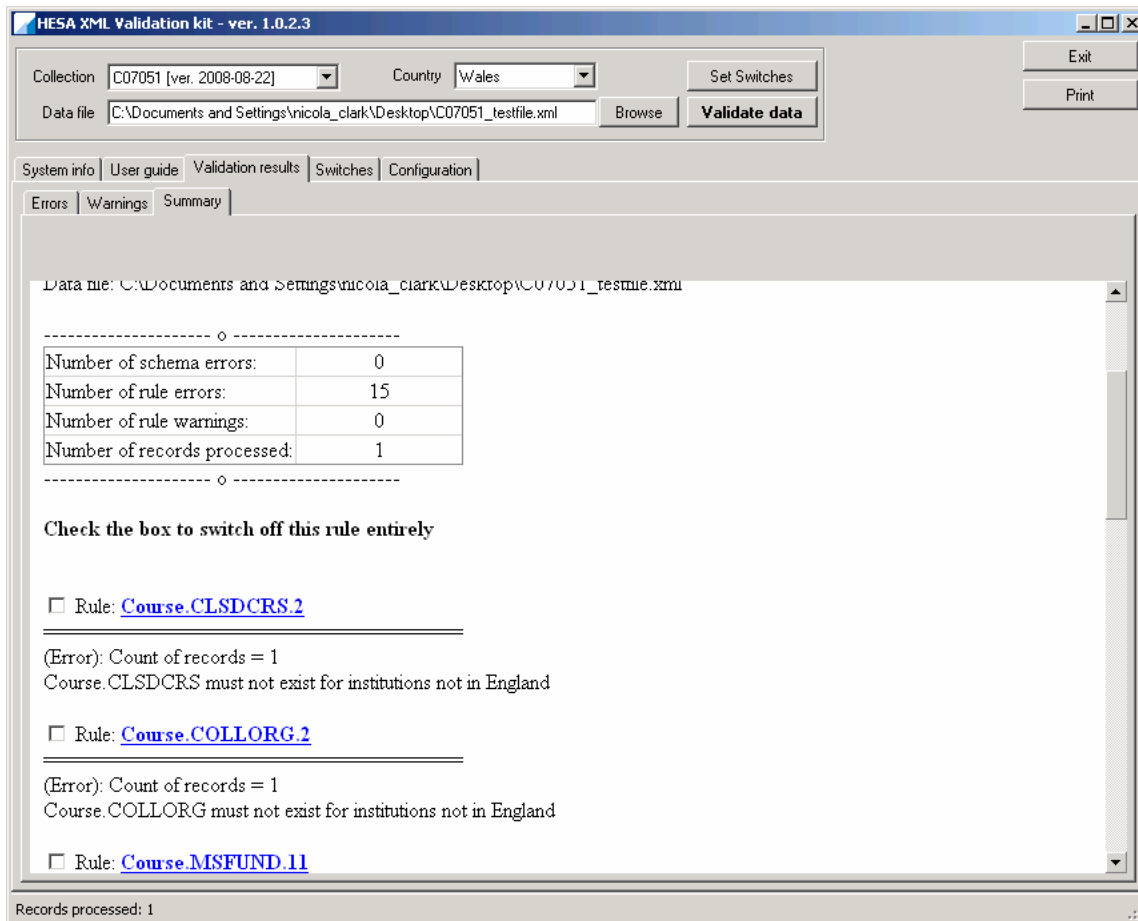
Warnings tab

The Warnings tab details individual records triggering warnings which should be considered to confirm that the coded data is genuine.



Summary tab

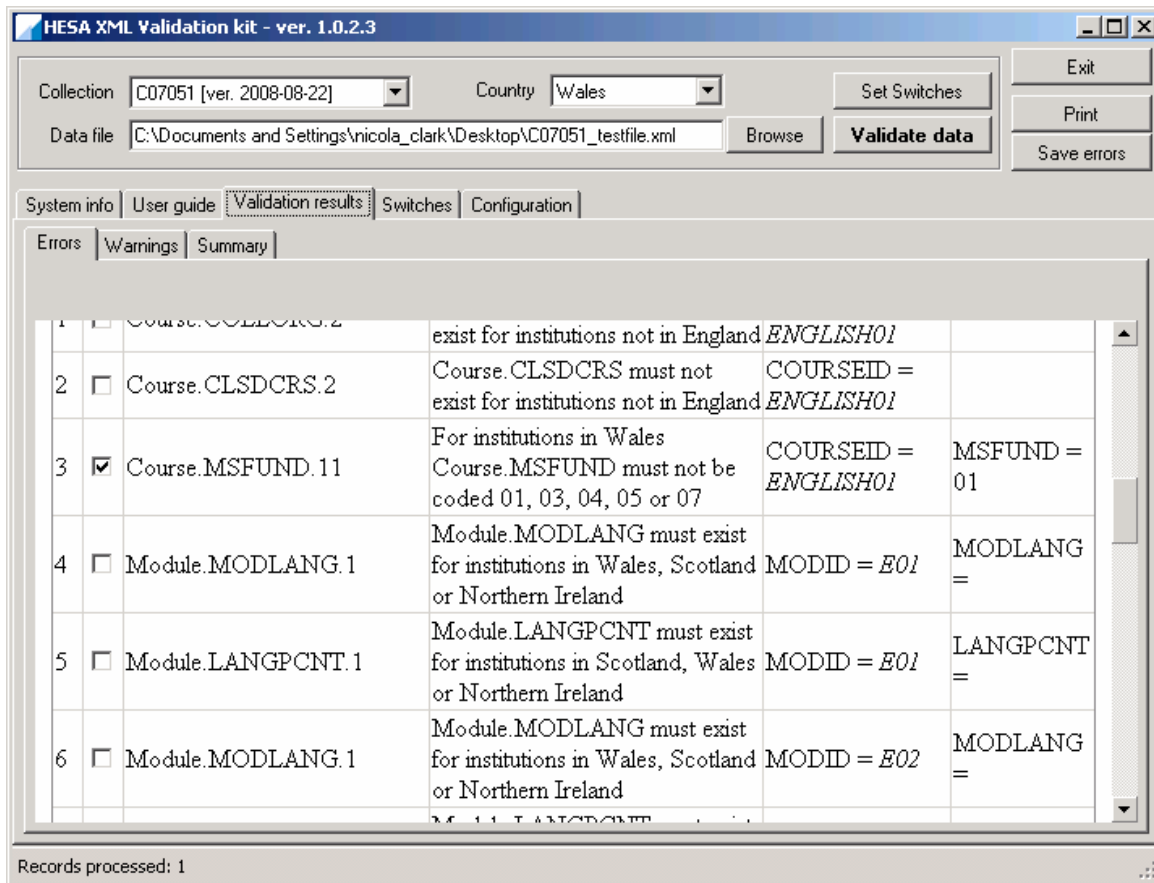
The Summary tab provides a summary total of the number of records failing a particular error or warning. The blue hyperlinks provide direct access to the relevant coding manual page to enable the user to cross-check the coding requirements of the field with which the validation rule is associated.



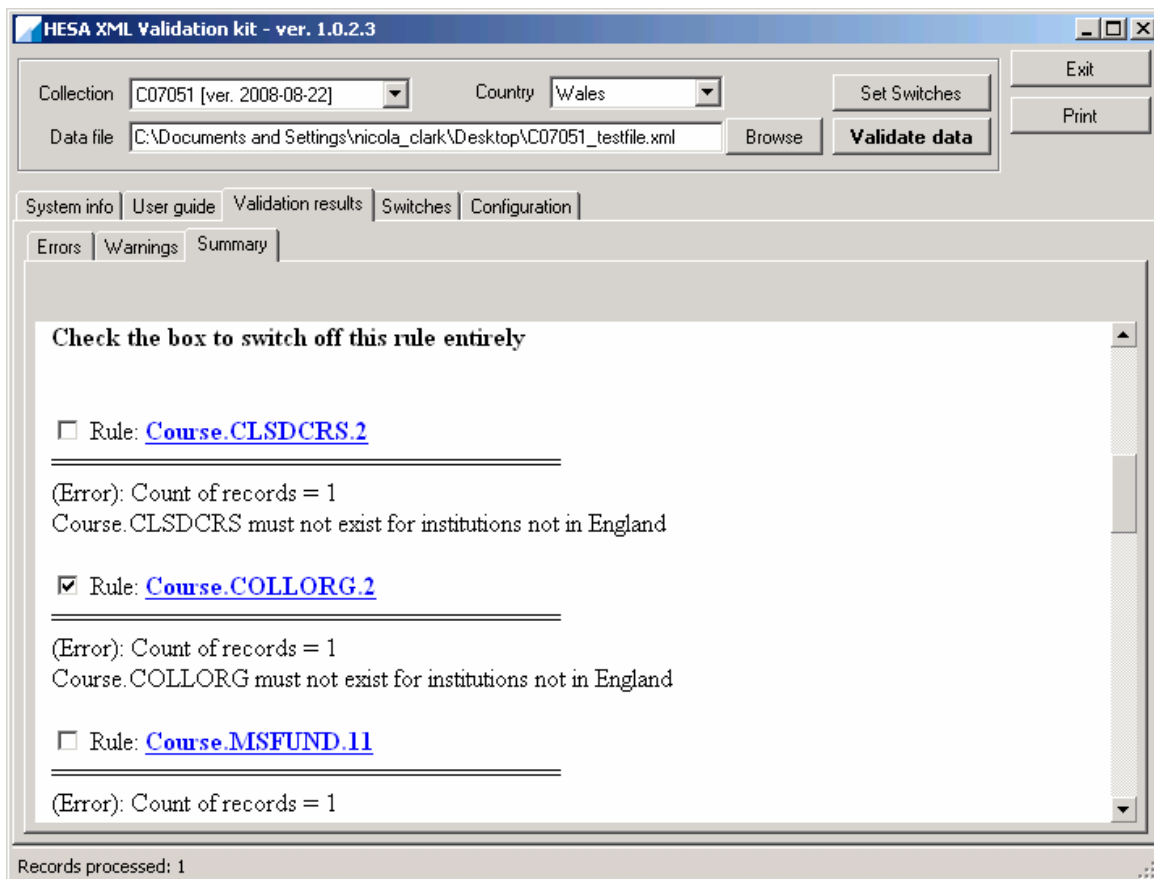
Setting switches

The validation kit provides users with two methods by which switches can be applied to deactivate validation rule. Note that only business rules can be switched off.

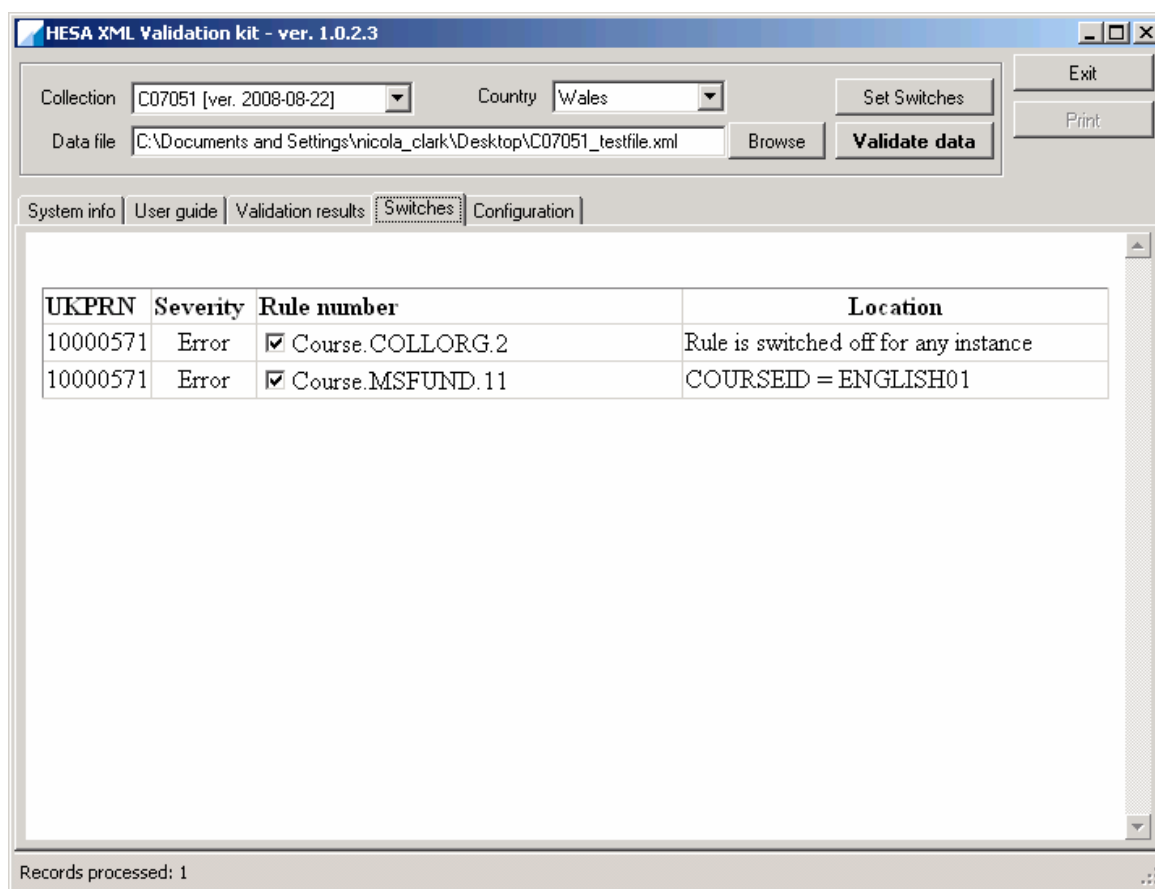
Through the Errors tab and Warnings tab switches can be applied to one particular validation rule for set records. To set a switch for a desired record click the tick box situated to left hand side of the error and then click the 'Set Switches' button.



Through the Summary tab switches can be applied across your file to prevent all records from failing. To set a switch against all records tick the check box to the left of the error text and click the 'Set switches' button.



To keep track of the switches which have been set view the 'Switches' tab. This tab logs all applied switches. From this tab, switches can be turned back off by de-selecting the switch you wish to re-activate and clicking the 'Set Switches' button.



Saving the error file

The 'Errors' and 'Warnings' tabs provide users with the option to save details of the validation errors into a tab delimited text file. This file can then be opened in Excel in order to provide a working document of listed errors.

Configuration

The 'Configuration' tab provides users with the option to specify certain functionality of the kit.

The screenshot shows the 'Configuration' tab of the HESA XML Validation kit. The window title is 'HESA XML Validation kit - ver. 1.0.2.3'. The interface includes a top navigation bar with tabs for 'System info', 'User guide', 'Validation results', 'Switches', and 'Configuration'. The 'Configuration' tab is active. The main area is divided into several sections: 'Rules, results and switches folders', 'Defaults (may impact performance)', and 'Institution data'. The 'Rules, results and switches folders' section contains fields for 'Folder with rules', 'Folder with results', 'File name for switches', and 'Folder for switches', each with a 'Browse' button. The 'Defaults (may impact performance)' section contains fields for 'Max number of errors', 'Size of a batch', and 'Max viewable file'. The 'Institution data' section contains a field for 'UKPRN'. There is also a 'Use Proxy' section with a checkbox and fields for 'URI:', 'Domain:', 'User name:', and 'Password:'. The status bar at the bottom reads 'Reading and validating data against the schema'.

Rules, results and switches folders

This section enables users to specify the location to which this data is stored on their local PC

Max number of errors

Once the number of errors (or warnings) has reached this value no further errors will be reported. Increasing this value may increase the time taken to produce the results. Entering a value of zero here will remove the limit and all errors will be reported.

Size of batch

The data is validated in batches of records, this is done to avoid loading the whole data file into memory which could cause your machine to run very slowly or crash.

Reducing the size of the batch will reduce the amount of memory required to process your file, however the kit will have to validate more batches and so may take more time.

Increasing the size of the batch will increase the amount of memory required to process your file but reduce the number of batches being processed, if your machine has sufficient memory this may improve the speed of validation.

This optimum batch size will vary between computers and will be influenced by the volume of data being passed through. You can try changing this value to gauge the effect of validating your data on your computer. HESA testing has found that values between 500 and 800 offer improved performance.

Max viewable file

If the result of a validation is very large it could cause memory problems when attempting to display them. This value provides a limit beyond which the results of the validation will not be shown in the kit (although it will be stored in the 'Folder with results' folder if you wish to view it in your web browser). If you encounter a message stating that the result file was too large you could modify this value. In any event you should be able to view the summary tab of the results page.

Use Proxy

Where your institution runs a firewall which prevents the kit from successfully communicating with the HESA servers, where the validation rules and updates reside, the proxy settings can be completed to facilitate the running of the kit. Tick 'Use Proxy' and set all the fields listed in that section to values applicable for your institutions configuration.