

QCS as a background service

1. Introduction

The QCS (Quality Check Software) is an application allowing to validate ENCR datasets. Starting from version 2.2.11, the QCS can be run also as a daemon-based application designed to validate datasets in the background.

This manual provides instructions about how to use the QCS application, including its features, configuration, and troubleshooting.

2. Getting Started

To start using the QCS application, you need to launch it with four arguments:

- `protocol_id`: the ID of the validation protocol
- `path_in`: the absolute path where the client deposits the input file(s)
- `path_out`: the absolute path where the QCS produces the output reports
- `polling_time`: the QCS polling interval for the `path_in` and `path_out` folders in seconds

For example, to launch the QCS application with the ENCR Incidence scenario (`protocol_ID = 11`), you can use the following command:

```
java -jar -Xmx2g jrc-qcs-2.2.11.jar -d 11 /home/user/input /home/user/output 3
```

3. How to Use the QCS Daemon

Once started, the QCS daemon executes an infinite loop, polling the `path_in` and `path_out` directories every `polling_time` seconds. To use the QCS daemon, follow these steps:

1. Create the dataset in the `path_in` directory (e.g., `/home/user/input/data.csv`)
2. Create an empty `flag_start` file in the `path_in` directory (e.g., `/home/user/input/flag_start`)
3. The QCS daemon will detect the `flag_start` file and verify that the `flag_running`, `flag_done`, and `flag_error` files are not present. If all conditions are met, the QCS daemon will start the validation process.

4. Validation Process

During the validation process, the QCS daemon performs the following steps:

- Selects the first CSV file in the `path_in` directory (e.g., `data.csv`)
- If more than one CSV file is found, the QCS daemon will write an error message to the `flag_error` file and delete the `flag_start` file
- Creates a `flag_running` file in the `path_in` directory and deletes the `flag_start` file
- Starts the actual validation process, producing all standard reports in its own internal directory (which must be different from the `path_out` directory)
- Copies the validation results to the `path_out` directory (e.g., `/home/user/output`)
- Deletes the `flag_running` file and writes a `flag_done` file to the `path_out` directory

5. Client Handling

Once the QCS daemon has completed the validation process (that is when the `flag_done` appears in the `path_out` directory), the client can handle the output reports in the `path_out` directory.

For example, the client can move or rename these reports, or remove from the `path_out` directory.

Remark: the client is fully responsible for the management of the `path_in` and `path_out` directories. This means that any issue due to a wrong management of these directories MUST be handled by the client (see "6. Error Handling").

6. Error Handling

If the QCS daemon encounters any error, during the orchestration or validation process, the QCS will write an error message to the `flag_error` file in the `path_in` or `path_out` directory, depending on the use case. The client MUST then handle the error accordingly.

Example: The QCS produces the `QCS-Incidence-Output.txt` file in the `path_out` directory. Later on, the user sets this file to read-only, without removing nor renaming it. The next time the QCS validates an input file, the QCS will try to copy the latest version of the `QCS-Incidence-Output.txt` file in the `path_out` directory, but this operation will fail, because the QCS cannot overwrite a read-only resource. In this case the content of the `path_out` directory would be incoherent, since the directory still contains the output reports of the *previous* input file (and not the output reports of the *current* input file). Therefore the `flag_error` file is created in the `path_out` directory, containing a short description of the problem. Then the QCS will continue polling both `path_in` and `path_out` directories, waiting for the client to fix the issue.

7. Bulk Mode

If the `polling_time` argument is set to 0, the QCS daemon will run in *bulk mode*. In this mode, the QCS daemon will validate all files in the `path_in` directory and produce output reports in the `path_out` directory. After completing the validation process, the QCS daemon will quit.

For example, to run the QCS daemon in bulk mode, you can use the following command:

```
java -jar -Xmx2g jrc-qcs-2.2.11.jar -d 11 /home/user/input /home/user/output
```

This will validate all files in the `/home/user/input` directory and produce all corresponding output reports in the `/home/user/output` directory.

Remark: This modality is oriented to be used in a automatic pipeline process, therefore only strictly necessary output reports are produced, which are:

- The `QCS-Incidence-Ouput-Summary.txt` report with an overview of the validation process
- The `QCS-Incidence-Ouput.csv` report with the detailed list of all errors and warnings