User guide:

Uploading *in situ* CWR passport data to EURISCO

Author:	Suman Kumar, Stephan Weise
Date:	2024-01-08
Version:	1.4

Content

1 Aim and scope

The aim of this document is to provide *in situ* National Focal Points (NFPs) with information about how to upload and update their *in situ* National Inventory (NI) datasets in EURISCO.

The general proceeding of an update is shown in the figure below. A National Focal Point formats data and then uploads them via the EURISCO *in situ* CWR intranet to a staging area. In the following steps, the data are checked for integrity and, after approval by the NFP, integrated in EURISCO. After final checks, the data are synchronised to a specific schema for the web frontend and are then available to the public.
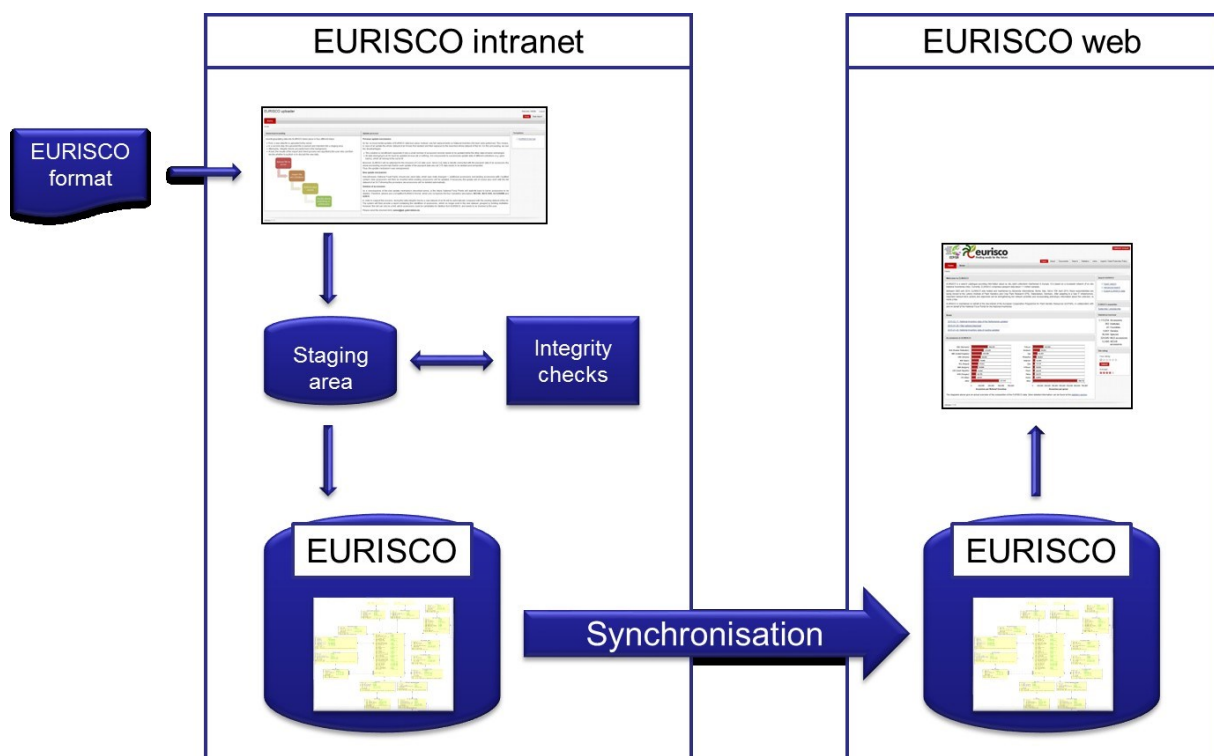
The individual steps are described in the following sections.
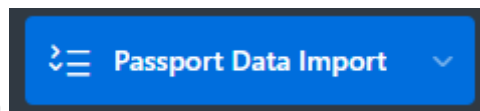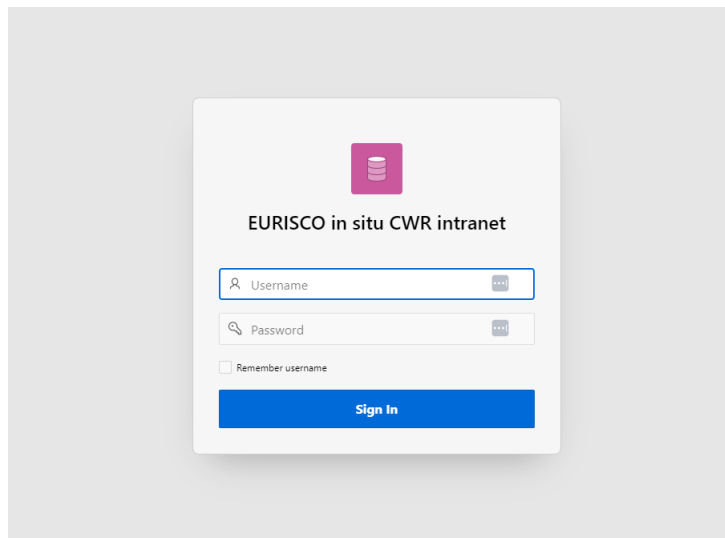
2 Procedure

2.1 DATA FORMATTING

In situ CWR passport data to be uploaded to EURISCO need to be formatted in accordance with the *in situ* CWR data standard, which is available at the documents section of the EURISCO web interface as well as in the EURISCO *in situ* CWR intranet.

2.2   DATA UPLOADING

Please login to the EURISCO *in situ* CWR intranet for uploading data.


Expand the navigation menu and click on
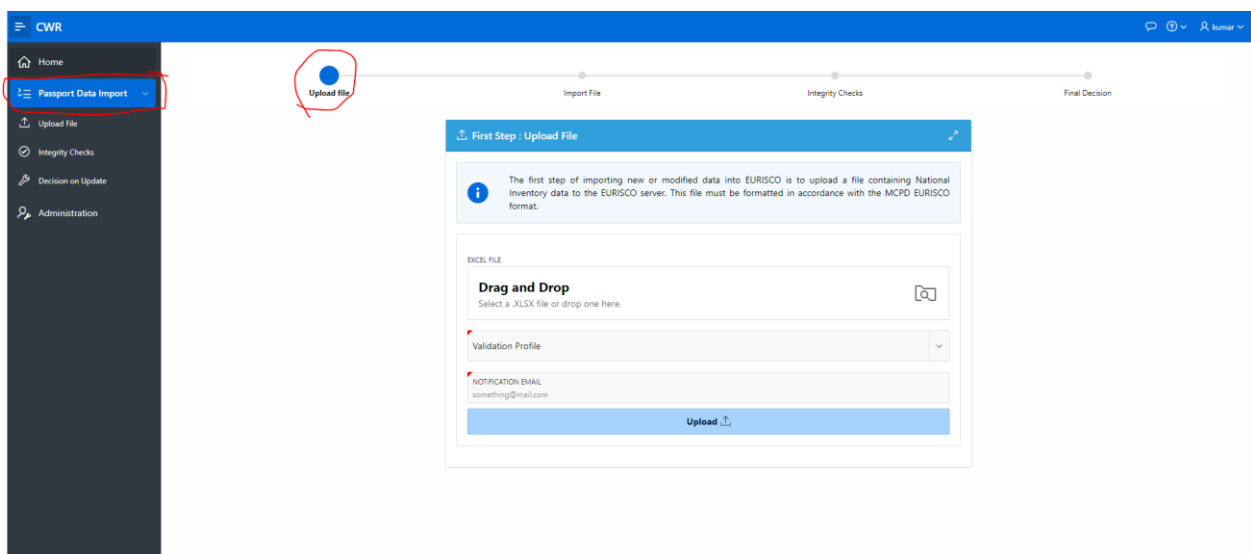
*The* in situ *CWR passport data uploader only supports files in the XLSX format (Excel 2007 and above).*

2.2.1   Steps to upload the Excel file

1. *Please choose the Excel file to be uploaded (as shown below)*
2. Select an appropriate validation profile from the select list:

   - **Errors abort session**
     If an error occurs, the whole import operation will be aborted.

   - **Records with errors are skipped**
     If an error occurs within a record, this record will be skipped. The remaining records will be imported.

   - **Descriptors with errors are skipped**
     If a descriptor value within a record contains an error, only this value will be skipped. The remaining descriptor values of the respective record will be imported. However, in case of an error within one of the mandatory descriptors, the whole record will be skipped.

3. Enter your email address for notification. Then, click on the "**Upload**" button.
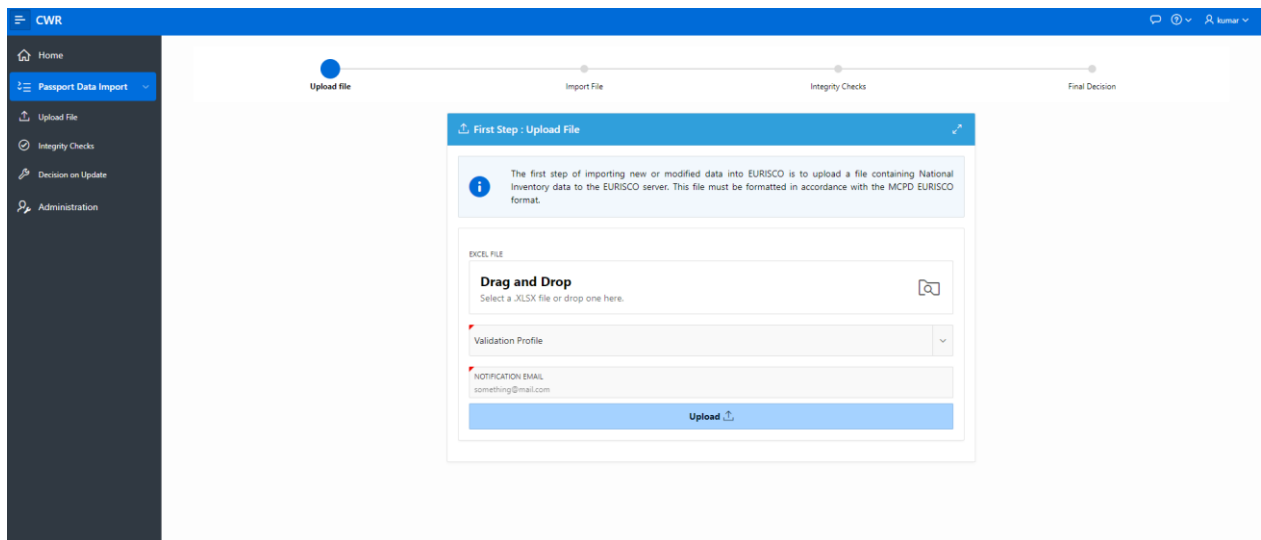
During upload, the first step is to check if all column headings conform to the *in situ* CWR data standard. If this is the case, a success message is displayed to the user after a completed upload.
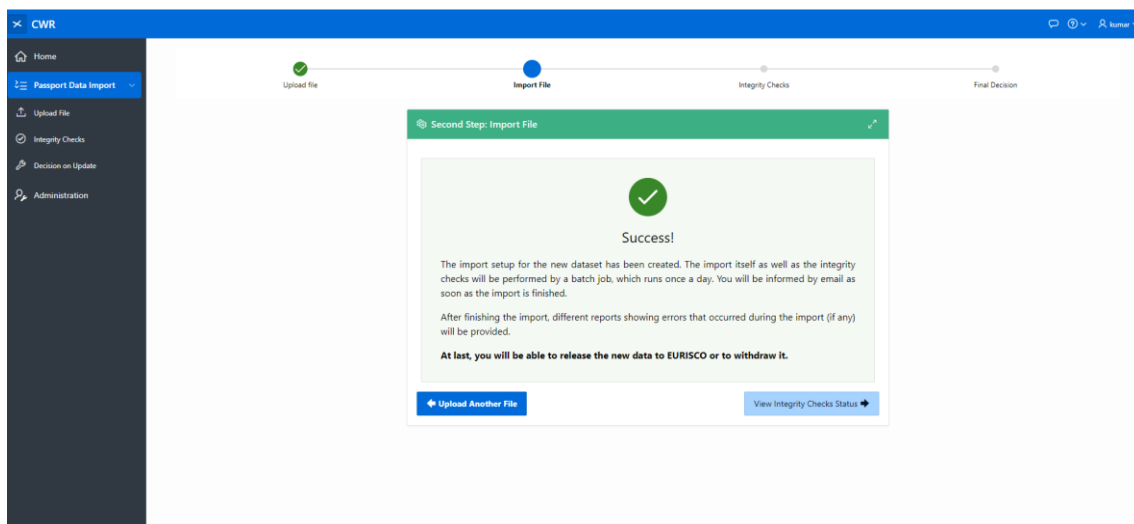
The uploaded Excel file is then parsed on the EURISCO server and the data is written to the EURISCO staging area. Subsequently, all necessary data integrity checks are carried out in the staging area.

This is done with a time delay by a batch job. You will be notified by email as soon as the integrity check results are available.

2.2.2 File validation errors

Allowed file extension is .xlsx:

The uploader filters and allows only to select .xlsx files.

Uploading the files without mandatory fields:

As per the _in situ_ CWR data standard, the uploaded Excel file must contain the mandatory fields NICODE, INSTCODE, ACCENUMB and GENUS. The upload process fails when the user submits the Excel file with missing mandatory fields.
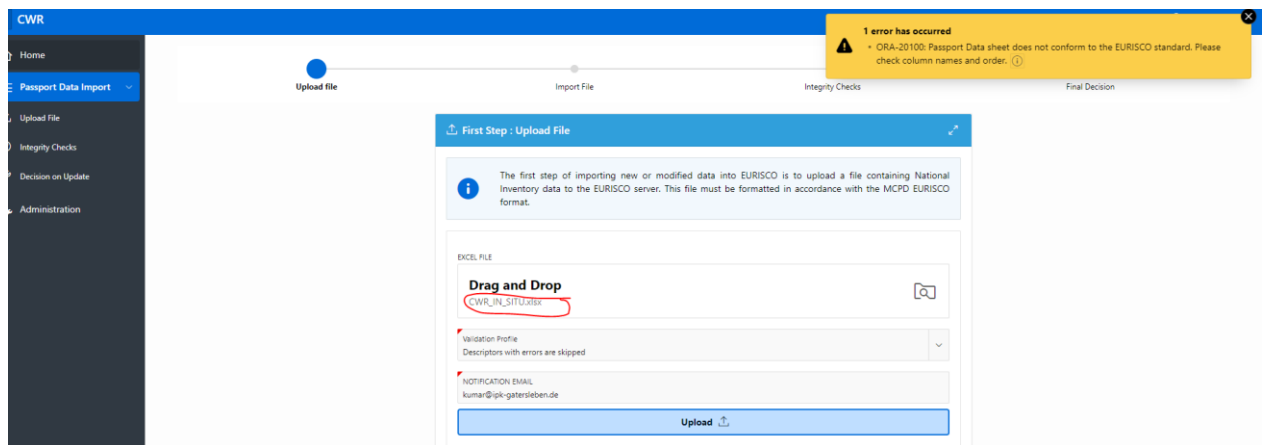
To fix the issue, please adhere to the file format available on the home page under **Templates**.
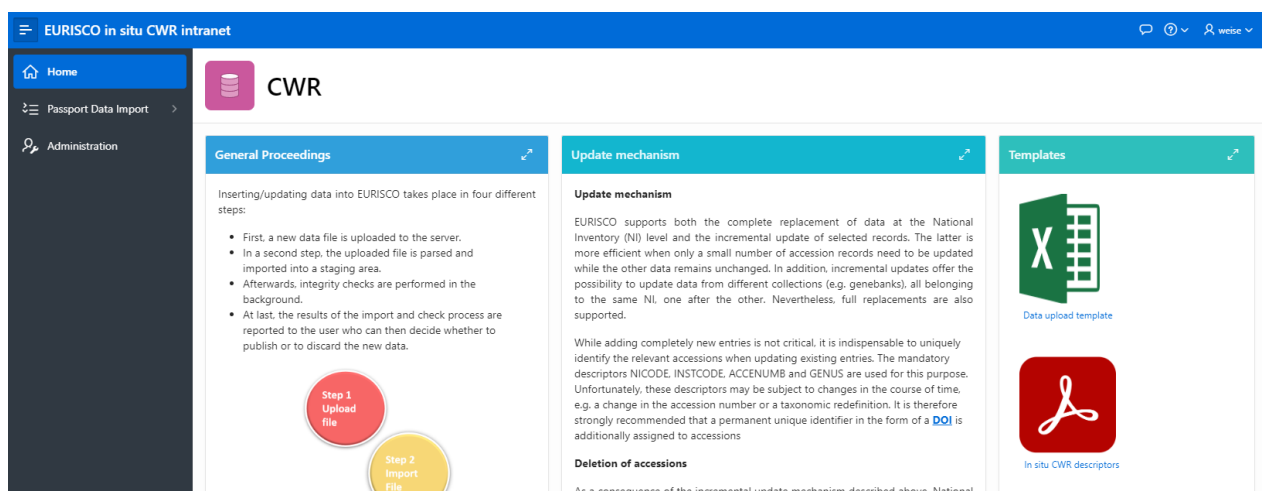
*Uploading the Excel file with incorrect header names:*

The file upload process fails when the user tries to upload the Excel file with incorrect header names, i.e. **HIST** instead of **HISTORIC** (error shown below).

To fix the issue, please don't change the template while editing.
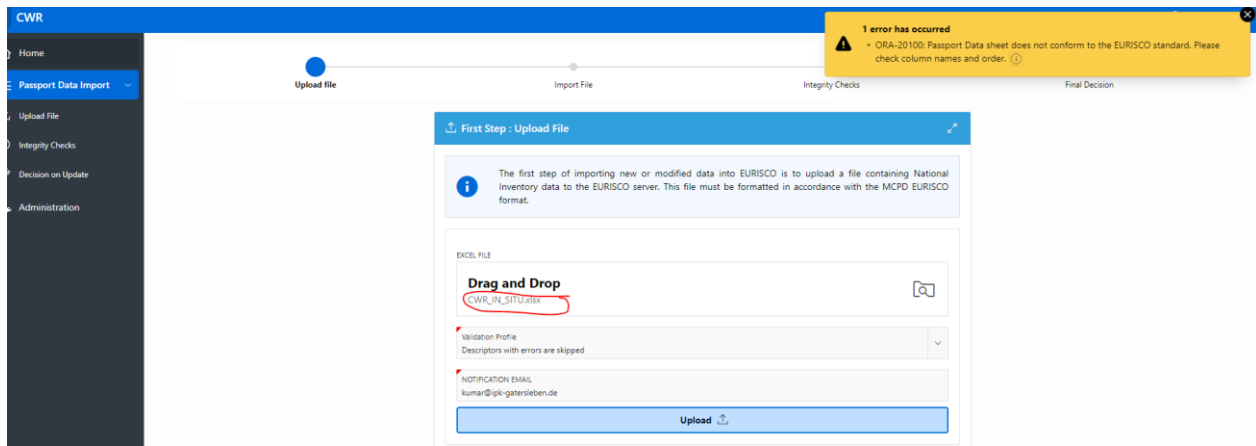
2.3   REVIEW INTEGRITY CHECK RESULTS

After you received a notification that the data integrity checks have been finished, please login to the EURISCO *in situ* CWR intranet again. Click on the menu item "Passport data import" and then on the sub-menu item "Integrity checks".
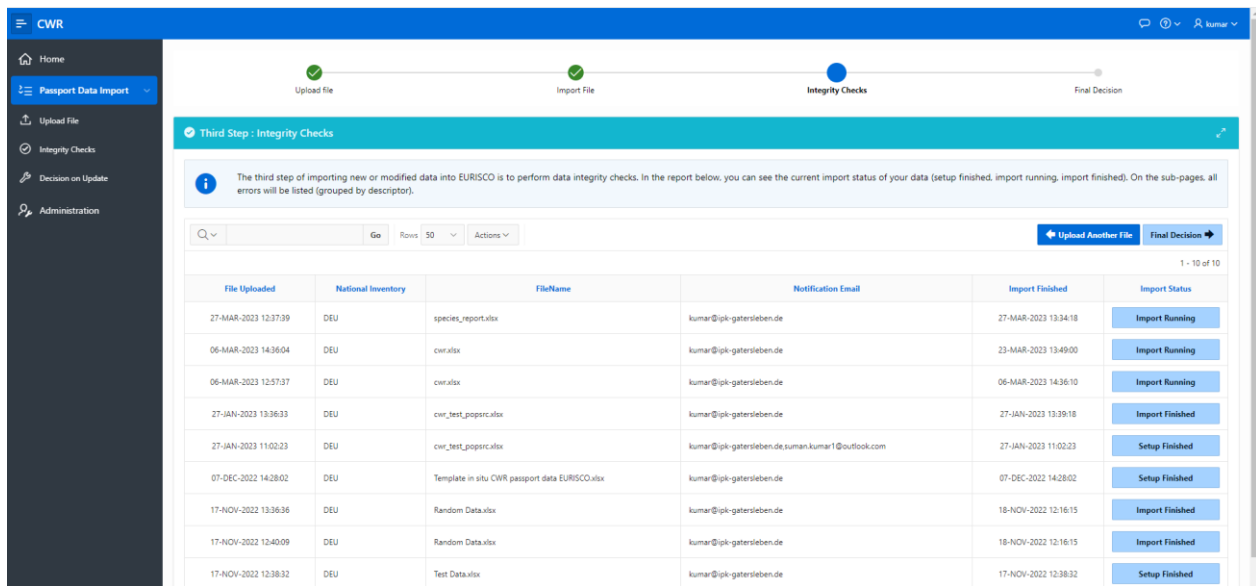
In the shown report, all uploaded datasets with their respective import status are listed. After the data integrity checks have been performed, the status will be "**Import finished**". The next steps now will be to review the errors that might have detected in your data and to decide whether to publish the data as they are (errors will be skipped according to your validation profile) or to rework them.

Please select a dataset and click on the "Import finished" button in order to see the error reports.

2.3.1 Check errors per descriptor

At the left side of the browser window, the number of errors is shown, grouped by descriptor. Click on the numbers next to the descriptor names to see all errors for the respective descriptors, e.g. STORAGE. In this example, several STORAGE values do not belong to pre-defined list of values.

Please go through all errors listed for the particular descriptors and check/correct the data file you uploaded. The *in situ* CWR data standard description is available from the start page of the intranet application as well as from the documents section of the EURISCO web application.

2.3.2 Check deletion candidates

Updating data in EURISCO does not mean that the whole dataset of a National Inventory will be replaced by the new dataset. Instead, incremental updates will be performed. This means that new accessions will be inserted while existing accessions will be updated. Thus, it is possible to successively update data of different collections (e.g. genebanks), which all belong to the same NI. If necessary, the update will of course also work with the full dataset of an NI. Following this procedure, no accessions will be deleted automatically.

In order to distinguish between new and existing accessions, the four mandatory descriptors NICODE, INSTCODE, GENUS, ACCENUMB are used. Together, they form an identifier, which should be unique within a National Inventory dataset.

Nevertheless, values of these four descriptors may change over time, e.g. a typo in the genus name may be corrected in the new dataset. In such a case, it is not possible for the import system to make a connection between the old record and the new record (with the corrected genus name). Thus, National Focal Points will explicitly have to name accessions to be deleted.

In order to support this process, during the data integrity checks a new dataset of an NI will be automatically compared with the existing dataset of this NI. The system will then provide a report containing the identifiers of accessions, which no longer exist in the new dataset, grouped by holding institution.

This list can only be a hint as to which accessions could be candidates for deletion from EURISCO, and needs to be checked by the NFP.

2.3.3 Check taxonomy mappings

Since the correctness of scientific plant names poses one of the most important challenges, the taxonomic terms provided to EURISCO are checked automatically against taxonomic repositories, currently against GRIN-Taxonomy and Mansfeld's World Database of Agricultural and Horticultural Crops, which represent opposite taxonomic schools.

Reports of matching/non-matching results of the provided terms and suggestions of closest existing matches are generated. These check results should only be considered as suggestions, but may help to improve the dataset.

2.3.4   Check AEGIS status changes

The AEGIS status of EURISCO accessions is checked automatically. According to the AEGIS principles, the reasons for de-flagging of accessions need to be documented. In case of de-flagging an accession previously flagged for AEGIS, the user will be asked for providing a change reason.

2.4 FINAL DECISION ABOUT UPDATE

After error checking, the final step is to decide whether the new data should be published in EURISCO or not. Please click on the sub-menu item "Decision on update".

Here, you will find a report showing all uploaded datasets that have been checked for data integrity. Please click on the "update/discard" button.

If your dataset needs revision, please click on "Discard data" and repeat the upload and check process with a reworked file. If your dataset does not contain errors or if you wish to skip the errors during the update, please click on "Update EURISCO data." Your data, which are still stored in a staging area, will then be used for updating EURISCO. Afterwards, the data will be synchronised to the EURISCO web frontend as described in section 1 (time lag; not in main business hours).