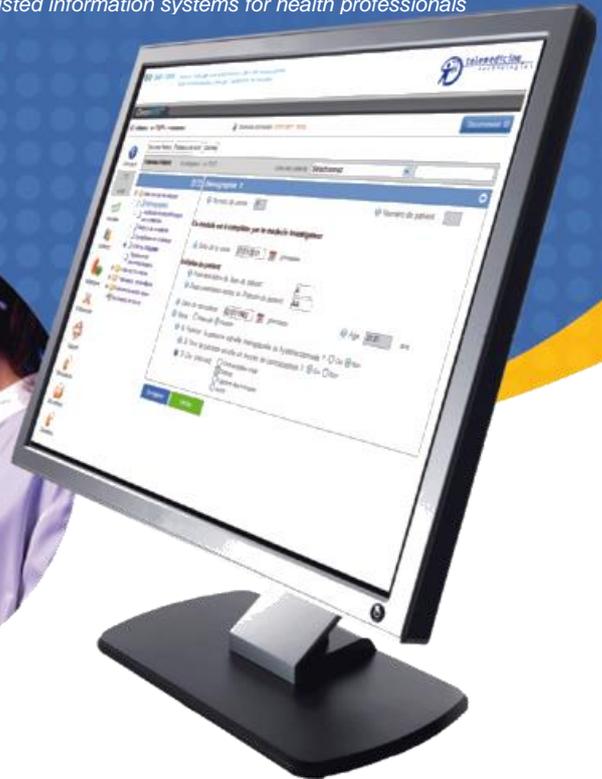




telemedicine  
technologies

Trusted information systems for health professionals



# CleanWEB

*The integrated solution for electronic management of clinical trials  
From designing the electronic case report form to freezing the database*

## Data Transfer Guide

---

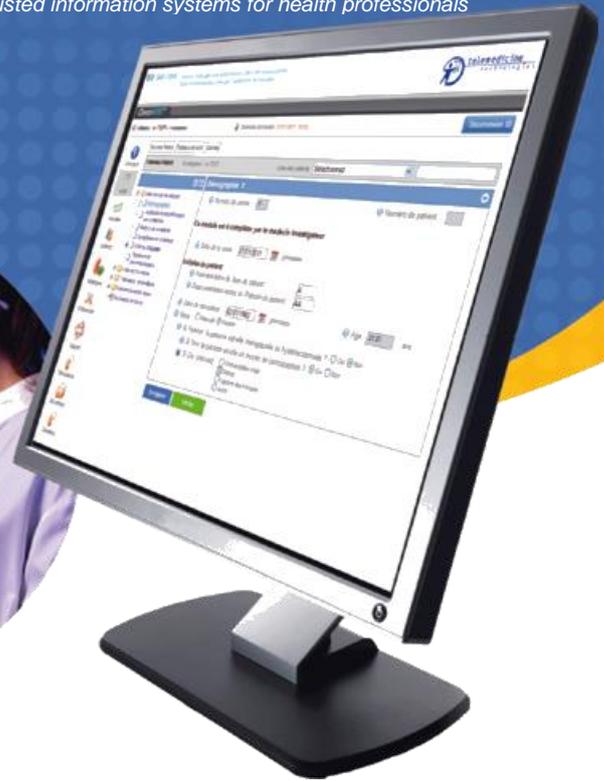
### AVIATOR

Référence	CW-206-USM
Version	2.0
Révision	0



telemedicine  
technologies

Trusted information systems for health professionals



## Contents

<b>1 - Purpose</b> .....	<b>4</b>
<b>2 - Principle</b> .....	<b>4</b>
<b>3 - Architecture</b> .....	<b>4</b>
<b>4 - Mapping</b> .....	<b>5</b>
<i>Common column headers in all data files</i> .....	5
<i>Numerical variables (Float, Integer)</i> .....	6
<i>Dates</i> .....	6
<i>Multiple choice questions</i> .....	6
<i>Text fields</i> .....	7
<i>Calculated variables</i> .....	7
<i>Patient identifying data</i> .....	7
<i>Module column</i> .....	7
<i>Specific columns in the file PATIENT_ANSI.csv</i> .....	8
<i>Specific columns in the file SURGICAL REPORT_ANSI.csv</i> .....	8
<i>Specific columns in the file POSTOPERATEDFOLLOWUP_ANSI.csv</i> .....	8
<b>5 - Import files</b> .....	<b>8</b>
<b>6 - Import validation</b> .....	<b>8</b>
<b>7 - Green light</b> .....	<b>8</b>
<b>8 - Upload frequency</b> .....	<b>8</b>

## 1 - Purpose

The purpose of this guide is to describe the steps to perform data transfer (batch upload) from a local database to the AVIATOR registry in CleanWEB™.

## 2 - Principle

- The participating center will receive an empty template dataset with the required column names.
- Then, the mapping between the institutional database and the AVIATOR database should be performed locally. Ideally, the mapping will be performed by a local data manager or a surgeon with data software skills.
- Completed datasets can be imported in the test environment of CleanWEB (this a copy of the production or live database).
- The application will perform logical checks on the dataset during the upload process. In case of errors - for example wrong variable names or incorrect formats - the system will generate an error report. This error report will be send to the investigator.
- When the upload went successful the content of the imported data must be validated in the test environment by the investigator.
- After green light from the investigator, the validated datasets will be uploaded to the production database.

The following documents will be provided:

- An empty template dataset (generated by the system through the extraction tool). The dataset contains four sub datasets in CSV format.
- The paper CRF (provided by the AVIATOR project) and the annotated CRF (provided by Telemedicine).
- The AVIATOR data dictionary with the variable names, codes, formats and definitions.

## 3 - Architecture

Data can be uploaded in CleanWeb only if it is organized exactly in the same manner as the architecture of the AVIATOR study.

This architecture includes:

- |                    |   |
|--------------------|---|
| Same file names:   | The sub datasets are named as follows: PATIENT_ANSI, SURGICAL REPORT_ANSI, POST OPERATED FOLLOW UP_ANSI, ECHO_1_ANSI<br>Be aware that the underscores only appear at the end (before ANSI), except for the ECHO file. |
| Same file format:  | All files must be saved as CSV (Comma Separated Value), in this case with a semicolon ; as delimiter (to mark the individual columns).  |
| Same architecture: | The content of the files should have the same architecture are the template files. Meaning, they should have exactly the same number, name and order of the columns within the files.                                 |

Same variable formats: The data within the columns should be in the correct format (such as date format, numeric values and text length for text variables).

## 4 - Mapping

The AVIATOR dataset contains of four template datasets in .csv format.

- PATIENT\_ANSI.csv containing patient data, like demography and status. Each row is a unique patient.
- SURGICAL REPORT\_ANSI.csv containing data of the Surgical Report module (preoperative, operative, complications and discharge). In the surgical report the first operation is recorded. This means that each row is a unique patient and unique first operation. AV related reoperation within the hospitalization of the first operation is registered in additional columns within the Surgical Report. The operative variables are copied. The same holds for Additional Clamp Sessions. These variables are copied into additional columns as well. Within the first hospitalization (Surgical Report module) it is possible to register 1 additional AV related reoperation and a maximum of 5 additional clamp sessions.
- POST OPERATED FOLLOW UP\_ANSI.csv containing data of the Follow-up module. Each row is one follow-up visit. As each patient can have several follow-up visits, each patient can have more than one row in the dataset. AV related reoperation within the follow-up period is registered in additional columns within the post-operative follow-up module. So be aware that the AV related reoperation should be registered in the correct dataset (so within the Surgical Report or the Post Operative Follow-Up file) according to the moment the reoperation was performed.
- ECHO\_1\_ANSI.csv containing data of the ECHO table. Each row is one echo performed in time. As each patient will have several echo's (preoperatively, intraoperatively before repair/replacement, intraoperatively after repair/replacement, postoperatively and during follow-up), each patient can have more than one row in the dataset.

The mapping has to be performed for each variable. When a question contains multiple answer options, the correct code has to be registered. Codes are listed in the data dictionary and in the annotated eCRF.

### Common column headers in all data files

The first nine columns are the same in all files. Most of them are facultative and can be left empty. They have to be completed with the same values for each unique patient, as described below:

#### Variable name

- STUDY_ID	Study name = AVIATOR in all files	<i>facultative</i>
- COUNTRY_ID	Country ISO code (2 characters)	<i>facultative</i>
- EXTRACTION_DATE	Extraction date	<i>facultative</i>
- SITE_ID	Center code	<u>mandatory</u>
- SUBJECT_ID	System reference	<i>facultative</i>
- SUBJECT_REF	COUNTRY_ID - SITE_ID - <i>sequential number</i> (e.g. FR-001-0001)	<u>mandatory</u>
- REF_COUNTRYISO	Same value as COUNTRY_ID	<i>facultative</i>
- REF_SITEREF	Same value as SITE_ID	<i>facultative</i>
- REF_NUMBER	This column should be left blank	<i>facultative</i>

## Numerical variables (Float, Integer)

Numerical variables are registered in 2 columns. The first column contains the value entered by the user (including values such as NA or N). The second column (with suffix \_V) should only be filled when the entered data is a numerical value. This column stays empty if the user indicated the item as unknown or not available. In that case the data cannot be saved in a numerical format but is saved as text.

For example, in the file SURGICALREPORT\_ANSI.csv the question 'Max diameter of aneurysm' (reference variable name: REFDIA) contains two columns: REFDIA and REFDIA\_V.

Floating versus Integer: when for a numerical value a floating format is required, the value may contain decimals. In case of an Integer, the number can not contain any decimals.

## Dates

Dates are registered in four columns: Day (DD), Month (MM), Year (YYYY) and Date (DD/MM/YYYY). The last column 'Date' is only filled when there are no unknown components in the Date. For example, if the user entered "ND/03/2011" the last column should be left empty. It will be the same if the user entered 03/14/2011, because the month "14" does not exist. When the date is known completely and registered correctly e.g. "12/07/2011" the last column has to be filled. In the file PATIENT\_ANSI.csv, four columns are created for the date of birth: DOB\_D, DOB\_M, DOB\_Y, DOB.



Dates have to be  $\geq 1900$  and  $\leq$  current year + 50 years. Otherwise the last column with the date format should be empty. Day and Month should be completed with a leading zero.

## Multiple choice questions

A multiple choice questions could have different formats: a question which requires one single answer or a question where multiple answers are possible. In the data dictionary you can look up which format is required.

There are three specific formats possible:

- combo-boxes:** this is a drop down list with one single answer option possible. The answer is registered as a numeric value and each code corresponds with an answer option. The combo box questions are registered in one column. For example, NYHA classification: 1 = class I, 2 = class II, 3 = class III and 4 = class IV.
- check-boxes:** these are online visible as square boxes, where more than one answer option is possible. Each answer option is registered in a separate column. The check box question contains the same number of columns as number of answer options. The column names are distinguishable by the suffix behind the variable name, which is:  $\_Cn$  (where  $n$  corresponds with the answer number). For example, in the cusp analysis, the checkbox LC (reference LCCA = Left Cusp within Cusp Analysis) has six possible answer options: 1 = normal, 2 = prolapse, 3 = fenestration, 4 = perforation, 5 = calcification, 6 = vegetation. So six columns have to be created in the SURGICAL REPORT\_AINSI file with the following column names: LCCA\_C1, LCCA\_C2, LCCA\_C3, LCCA\_C4, LCCA\_C5, LCCA\_C6. These columns have to be filled with 1 if the checkbox has been checked (if the answer = Yes) and it should be left empty if not. Note: not 1 and 0, but 1 and empty.

**radio-buttons:** these are online visible as round circles, where one single answer option possible. For example, gender: 1 = male and 2 = female. Another common one is the Yes/No question.

The specific codes for the answer options are mentioned in the annotated eCRF and the data dictionary.

### Text fields

For text fields a maximum length of characters is defined. The maximum length will be checked during the import. See the data dictionary for the maximum values.

### Calculated variables

The registry contains some calculated variables, based on information entered elsewhere within the CRF. These variables will be calculated automatically by the application during the upload process. This means that these columns could be left empty for the batch upload. It is important not to delete the columns from the template dataset as the structure of the import file should not be altered.

### Patient identifying data

The import of patient identifying data is possible. These are: the local hospital patient ID, last name and first name of the patient. Identifying data will be encrypted and saved in a separate database (and server) within CleanWeb. Each participating center can decide for itself if it wants to register identifying data or not. The local Medical Ethical Committee should be informed.

### Module column

Each file contains a specific Module column (column J). This column has to be completed as follows:

File name	Column name	Value
PATIENT_ANSI.csv	MODULE_PATIENT	PATIENT
SURGICAL REPORT_ANSI.csv	MODULE_SURGICAL REPORT	SURGICAL REPORT
POST OPERATED FOLLOW UP_ANSI.csv	MODULE_POST OPERATED FOLLOW_UP	POST OPERATED FOLLOW-UP_ $n+1$
ECHO_1_ANSI.csv	MODULE_ECHO	ECHO_ $n+1$

For the PATIENT and SURGICAL REPORT file the values in the MODULE columns are the same for each row. For the POST OPERATED FOLLOW UP and ECHO file the value in the MODULE column depends on the number of occurrences for each individual patient (N+1).

For example: the row that contains the first follow-up visit, the following should be registered in the MODULE\_POST OPERATED FOLLOW\_UP column:

The next Follow-up visit will be registered as thereafter

POST OPERATED FOLLOW-UP\_1  
POST OPERATED FOLLOW-UP\_2  
POST OPERATED FOLLOW-UP\_3  
etc.

*! Be aware of the underscores usage. These should not be placed between each word, but only after MODULE and after the repetitive occurrence of multiple rows in the follow-up and echo file.*

#### **Specific columns in the file PATIENT\_ANSI.csv**

COUNTRYISO (column K) and SITEREF (column L) are identical to COUNTRY\_ID and SITE\_ID. They are not visible online but are used to build the AVIATOR\_ID. AVIATOR\_ID should be the same as SUBJECT\_REF.

#### **Specific columns in the file SURGICAL REPORT\_ANSI.csv**

CREAT and CREAT\_V: both variable should be registered in mg/dL. During the upload process the  $\mu\text{mol/L}$  values will be calculated automatically and will be visible online. If the local database contains the  $\mu\text{mol/L}$  values only, divide these values through 88,4 to get the required mg/dL values ( $1 \text{ mg/dL} = 88,4 \mu\text{mol/L}$ ).

#### **Specific columns in the file POSTOPERATEDFOLLOWUP\_ANSI.csv**

YEAR\_FIRSTSTOP and MONTH\_FIRSTSTOP: these two items will be calculated by the application after upload, so these can be left empty.

## **5 - Import files**

If the four files are completed, please send these in one archive file (zip) to your contact person of the batch upload.

## **6 - Import validation**

- First check if all patients were uploaded in the test environment. The patient list is available in the dashboard. The list can be downloaded to Excel.
- Check if the AVIATOR\_ID's are generated correctly by the system.
- Check if the source data are imported correctly. For example, by selecting a random sample of patients and go through the whole CRF and check each variable. In addition, select some important variables, like mortality, numeric measures or major complications, and check these for additional patients. Select complex cases. For example, patients with more than one clamp session. Select patients with a long follow-up duration and patients with reoperations. Check if the multiple occurrences were uploaded correctly.

## **7 - Green light**

After green light from the investigator, the validated datasets will be uploaded from the test environment to the production database.

## **8 - Upload frequency**

Decide together with your contact person of the Aviator project how to perform the upload for future patients and the frequency of uploading follow-up data for current patients.