General European OMCL Network (GEON) QUALITY MANAGEMENT DOCUMENT

PA/PH/OMCL (08) 87 R6

VALIDATION OF COMPUTERISED SYSTEMS
ANNEX 1 – VALIDATION OF EXCEL SPREADSHEETS

Full document title and reference	Validation of Computerised Systems Annex 1 – Validation of Excel Spreadsheets PA/PH/OMCL (08) 87 R6
Document type	Guideline
Legislative basis	-
Date of first adoption	May 2009
Date of original entry into force	July 2009
Date of entry into force of revised document	August 2018
Previous titles/other references / last valid version	Validation of Computerised Systems Annex 1: Validation of computerised calculation systems: example of validation of in-house software PA/PH/OMCL (08) 87 2R
Custodian Organisation	The present document was elaborated by the OMCL Network / EDQM of the Council of Europe
Concerned Network	GEON

**ANNEX 1 OF THE OMCL NETWORK GUIDELINE
"VALIDATION OF COMPUTERISED SYSTEMS"**

VALIDATION OF EXCEL SPREADSHEETS

Note: Mandatory requirements in this guideline and its annexes are defined using the terms «shall» or «must». The use of «should» indicates a recommendation. For these parts of the text other appropriately justified approaches are acceptable. The term «can» indicates a possibility or an example with non-binding character.

1. INTRODUCTION

This is the 1st Annex of the core document "Validation of Computerised Systems", and it should be used in combination with the latter when planning, performing and documenting the validation process of Excel® spreadsheets used for the processing of laboratory data.

This Annex presents an example of Excel spreadsheet validation, which should be used in combination with the general requirements and recommendations given in the core document.

2. INSTALLATION AND SECURITY

To guarantee that only the latest validated version of the spreadsheet is being used and to maintain the validated state of the spreadsheet, all validated Excel spreadsheets should be stored with read-only access rights for the end users (e.g., on a protected network share). Only responsible persons should have write access to the network share.
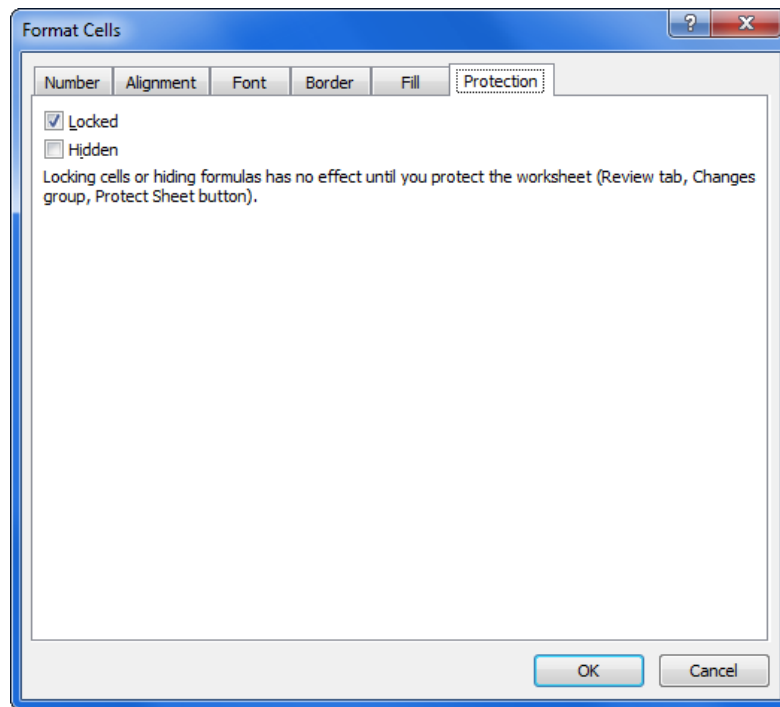
End users should have no right to modify a validated spreadsheet, add a non-validated spreadsheet to the share, or save data on the share. End users should only have the right to fill in the (permitted) cells and to print the data or save a copy to a data repository if needed.

Installation shall be documented, e.g. in the validation file, in a system log book or on a QA form. The name of the spreadsheet, unique identification, localisation, and the person responsible for the spreadsheet shall be documented. The records shall also include verification, regular verification and other issues such as updates or any problem encountered. Verification is completed after installation and recorded.
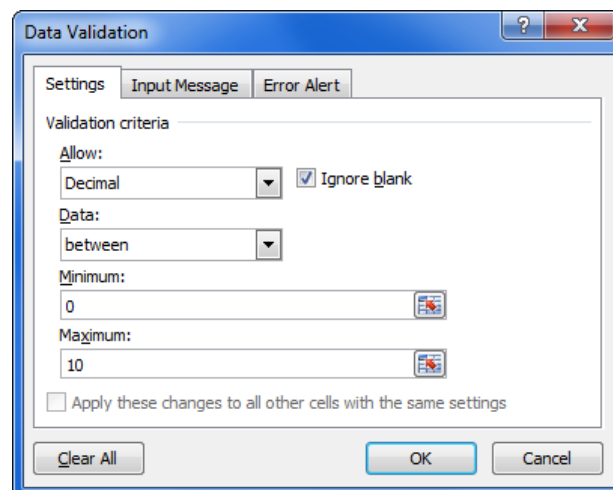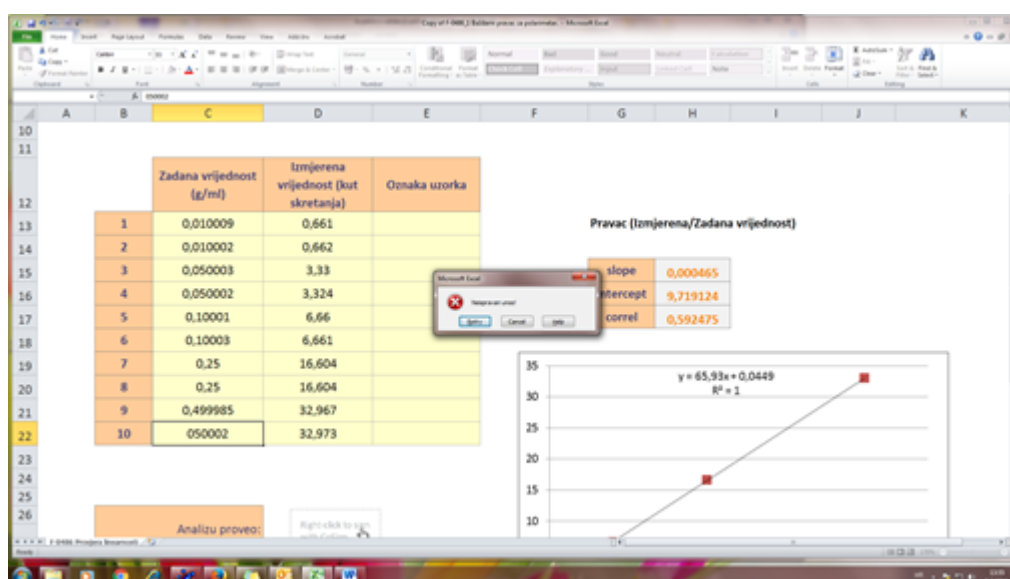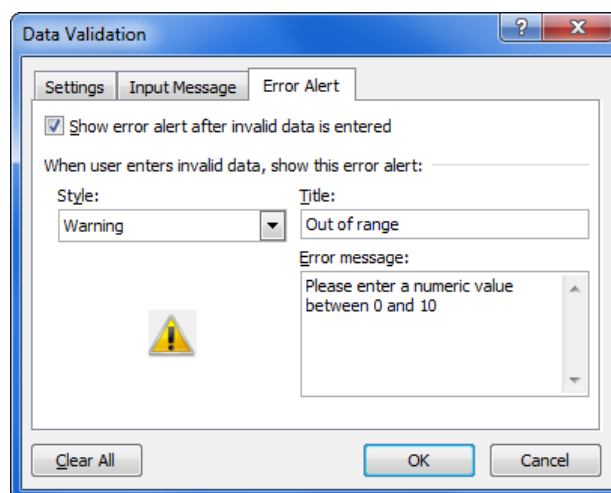

3. GOOD PRACTICES

When setting up a new spreadsheet, following the good practices below will reduce the risk of accidental modifications of the template and erroneous data input:

- All calculating cells shall be locked (Format Cells > Protection > Locked) in order to protect cells containing calculations against unintended modification, except those used for data input.
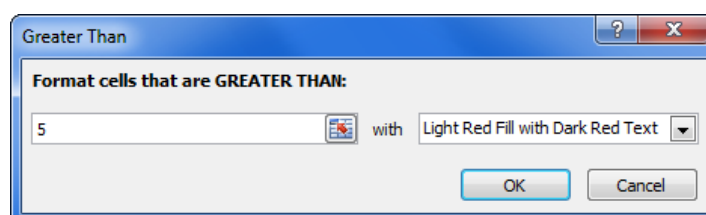
- Cells used for data input can be identified by a specific colour.

- Data validation rules (Data tab > Data Validation) can be applied to data input cells to prevent the introduction of aberrant values. Input messages and Error alert messages can be used to inform the end user of the expected data type and acceptable range.

- Cells used for presenting the results of the calculations (output) can be identified by a specific colour. When the results are tested against acceptance criteria it is recommended using conditional formatting (Home tab > Conditional Formatting) to highlight out-of-specifications results.
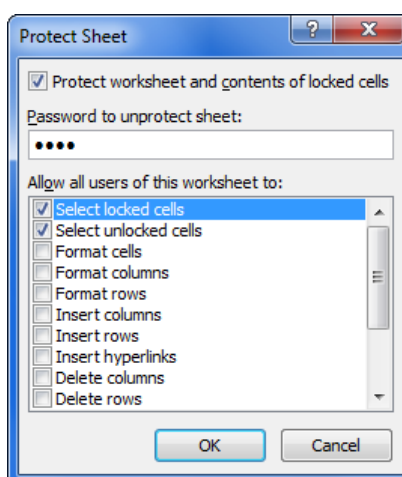

- The name of the operator responsible for data entry, and the date and time of data entry should be recorded in dedicated input cells or the spreadsheet is printed, signed and dated after calculation.

- File path, spreadsheet filename and MS Excel® version number can be displayed within the print area of the spreadsheet. The Excel functions '=CELL("filename")' '=INFO("RELEASE")' can be used to display the path, filename, active sheet and the version number of MS Excel® in use.
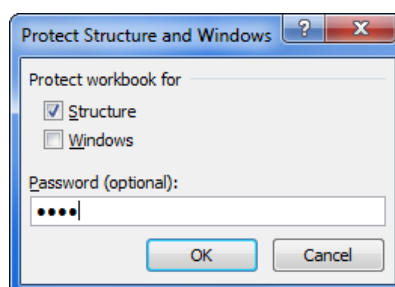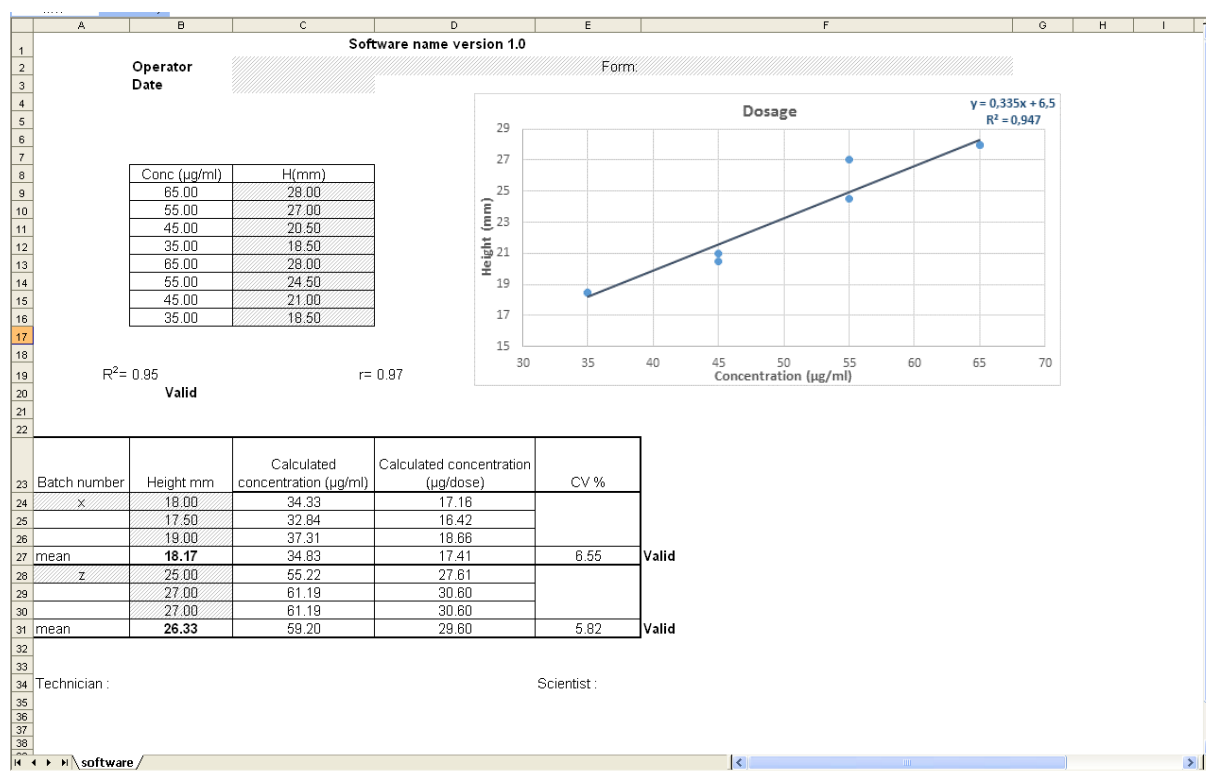
- Password protection is recommended for all cells containing calculations (Review tab > Protect Sheet), with only the default options checked. The same password can be used for all sheets and can be documented in the validation file. The sheet protection password should not be communicated to the end users.

- After protecting each sheet, the workbook structure should also be password protected (Review tab > Protect Workbook). The same password can be used as the one for sheet protection.

An example of a spreadsheet used to calculate a vaccine titration is shown on the image below. From results obtained for a reference product (height measured at 4 concentrations), a calibration curve and its formula are provided. Both of them are needed to calculate the concentrations corresponding to the height measured for the tested vaccine.

In the image, grey cells are filled with numerical data from experimentation and are the only ones that can be changed by the operator. All other cells are locked. No more than one cell from the calibration range can be empty; all cells for vaccines must be filled to guarantee proper use.

4. VALIDATION STAGES

4.1. Documentation of the spreadsheet

There should be a general description of the spreadsheet explaining its purpose, general layout, input types and data validation rules if required (some spreadsheet might be self-explaining). This description can be documented in the spreadsheet itself (e.g. in a dedicated sheet), in a SOP or in the validation file.

Next to the general description, a full print-out of the spreadsheet where all formulas are shown (Formulas tab > Show Formulas) should be kept in the validation file.

When VBA[1] macros are used, the VBA code should also be printed and kept in the validation file.

If matrix-formulas (array-formulas) are used, this must be indicated. An individual printout of each matrix formula is necessary.

---

[1] Visual Basic for Applications

All print-outs shall clearly identify the spreadsheet name or identification and version number. When a new version of the spreadsheet is being validated, a summary of the changes since the previous version should be given.

The version of Microsoft Excel used for the creation and validation of the spreadsheet should be traceable (either by the documentation of the spreadsheet or by the change log of the IT department), and any known incompatibilities with older or newer versions should be documented.

The documentation of the spreadsheet can be considered as the URS.

In order to properly document the spreadsheet, formulas shall be printed and entered into the validation document (see example below).

4.2. Validation of the calculations of the spreadsheet

All calculations are to be verified with a system completely independent from the self-developed spreadsheet. One validation method is to compare the results obtained by the spreadsheet with results obtained by commercial software or with a calculator, using the same dataset as input. Another validation method is to compare the results obtained by the spreadsheet with published reference data (e.g. physicochemical data of substances).

If the spreadsheet will be used on computers running different versions of Excel it is required to perform the validation of the functionality using each of those different versions as some newer Excel functions are not retro-compatible with older versions of Excel.

4.2.1. Validation of the calculations by using commercial software or published data

A dataset as close to real values as possible must be chosen. Excel calculations are compared to the results given by commercial software or by published data, which are considered as validated (see example in the image below). The commercial software provides the coefficient of correlation, R^2 and the coefficients of the calibration curve.

```
Regression Analysis - Linear model: Y = a + b*X
-----------------------------------------------------------------------
Dependent variable: H
Independent variable: Conc
-----------------------------------------------------------------------
                             Standard          T
Parameter        Estimate      Error       Statistic       P-Value
-----------------------------------------------------------------------
Intercept          6.5         1.6569       3.92299         0.0078
Slope              0.335       0.0323393    10.3589         0.0000
-----------------------------------------------------------------------


                      Analysis of Variance
-----------------------------------------------------------------------
Source          Sum of Squares   Df  Mean Square    F-Ratio      P-Value
-----------------------------------------------------------------------
Model              112.225        1     112.225      107.31       0.0000
Residual             6.275        6       1.04583
-----------------------------------------------------------------------
Total (Corr.)      118.5          7

Correlation Coefficient = 0.973163
R-squared = 94.7046 percent
Standard Error of Est. = 1.02266


The StatAdvisor
---------------
   The output shows the results of fitting a linear model to describe
the relationship between H and Conc.  The equation of the fitted model
is

   H = 6.5 + 0.335*Conc

Since the P-value in the ANOVA table is less than 0.01, there is a
statistically significant relationship between H and Conc at the 99%
confidence level.

   The R-Squared statistic indicates that the model as fitted explains
94.7046% of the variability in H.  The correlation coefficient equals
0.973163, indicating a relatively strong relationship between the
variables.  The standard error of the estimate shows the standard
deviation of the residuals to be 1.02266.  This value can be used to
construct prediction limits for new observations by selecting the
Forecasts option from the text menu.
```

If no discrepancy occurs, the validation of this part of the calculation is considered as fulfilled. If a discrepancy is observed, a check and revision of the formulas must be performed (and the whole validation re-performed).

4.2.2. Validation of the calculations with a calculator (manual calculation)

Using the printed formulas from the spreadsheet, all concentrations are calculated using a calculator (see next image) and compared with the results given by the spreadsheet.

As an alternative, the PC calculator can be used and documented in screen shoots, as in the image below.

If no discrepancy occurs, the validation of this part of the calculation is considered fulfilled. If a discrepancy is observed, both the revision of the formulas and the manual calculations should be repeated (and the whole validation re-performed).

Moreover, calculations in paragraph 4.2.1 and 4.2.2 should be re-performed with other datasets including exceptional situations, for example: OOS results, missing data, or nonsense data. Calculations should also be validated under these conditions, as applicable (data not shown).

4.2.3. Validation of the protections

The following points shall be verified and documented:

- Access rights to the spreadsheet (e.g. on the network share) are correct: the file cannot be modified or deleted by users.
- The different sheets within the spreadsheet are properly protected: only input cells can be edited, all other cells are locked.
- A password (if applicable) is needed to remove sheet protection and workbook protection.

At this stage, the spreadsheet is considered as validated and its status is issued and filed.

5. REGULAR VERIFICATION OF THE SPREADSHEET

Regularly, in a risk-based approach an OMCL should define an appropriate frequency of regular verification of an existing spreadsheet. After every change performed in the soft- or hardware configuration, the spreadsheet should be verified to ensure that its validated state is maintained. A known dataset is used and the results are compared to the standard one.

In order to help the operator, verification instructions containing the information required should be available.

Each verification is registered, with the following information: date of operation, intervention (i.e. verification), comments, and operator's signature. Results from the verification should be kept in the validation file or system documentation.