

# Spreadsheet standards

Department for Education guidance

2021-08-24



# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	Purpose . . . . .	5
1.2	Design principles and good practice . . . . .	5
<b>2</b>	<b>Workbook, worksheets and formula design</b>	<b>7</b>
2.1	Introduction . . . . .	7
2.2	Workbook design . . . . .	7
2.3	Worksheet design . . . . .	9
2.4	Formula design . . . . .	10
<b>3</b>	<b>Good practice for Excel-based modelling</b>	<b>13</b>
3.1	Introduction . . . . .	13
3.2	Good practice in Excel . . . . .	13
3.3	Good practice spreadsheets . . . . .	15
3.4	Useful links . . . . .	15



# Chapter 1

## Introduction

### 1.1 Purpose

This document sets out the DfE standards for **all** modellers to follow when designing spreadsheet models. The standards apply to *all* spreadsheet models built, even quick and simple ones.

Presentation and design must be to high standards: tidy, logical, well-presented and well-laid out spreadsheets are easier to build correctly first time; easier for others to follow; and easier to audit and quality assure. A useful rule of thumb to follow here is that the effort invested in the clarity and layout should be commensurate to the complexity and/or importance of the model.

Modellers should be aiming for a standard that means another analyst coming to the model completely fresh can understand what the spreadsheet is doing within 30-60 minutes. We must **all** remember that someone may need to revisit the model at some point in the future. This may be for quality assurance purposes, where following these standards will reduce the number of questions for the model builder. Or perhaps the initial modeller is no longer involved in the work, with an increase in flexible resourcing this will become more prevalent.

This document works alongside the Department's Quality Assurance Guidance.

### 1.2 Design principles and good practice

Chapter 2 sets out the rules for spreadsheet modelling, organised into three main categories: workbook design; worksheet design; and formula design. Modellers should adhere to these rules, and any decision not to comply should be a conscious one and one that can be explained and justified.

In addition to the rules, we should acknowledge that individual modellers will have their own preferences and style, and different types of analysis or models will warrant different approaches. Therefore, Chapter 3 sets out additional good practice guidance, together with links to external guidance on spreadsheet modelling. The chapter also suggests the sharing of exemplar spreadsheets through the quality assurance area on the DfE intranet.

## Chapter 2

# Workbook, worksheets and formula design

### 2.1 Introduction

Chapter 2 sets out the rules for spreadsheet modelling, organised into three main categories:

- workbook design: rules relating to the high-level layout of your model;
- worksheet design: rules relating to the layout of worksheets within your model; and
- formula design: principles to ensure formula simplicity and formula clarity.

### 2.2 Workbook design

**Group and separate worksheets by type:** You should have four groups/types of worksheets:

- control (e.g. containing check sheets, version control, table of contents, date created);
- foundation (e.g. containing the model's inputs);
- workings (e.g. containing the model's calculations); and
- presentation (containing the outputs of the model).

This separation offers numerous advantages; for example, ‘workings’ sheets should not be made ‘pretty’ at the expense of calculation clarity – this can be achieved on ‘presentation’ sheets which are designed to be seen by the commissioning policy customer.

For smaller spreadsheet models, fewer worksheets can be used (e.g. merge the ‘control’ and ‘foundation’ sheets), as long as the layout is the same as a large spreadsheet model with multiple tabs.

**All models must have a control sheet (or sheets):** as a minimum, these should describe the purpose of the model, give the name of the analyst and date of creation, include a table of contents (with description of each sheet), provide details of the quality assurance undertaken, provide a key for any colour coding used, and list model assumptions and caveats.

**All input assumptions must have a recorded basis:** as a minimum, a full list of assumptions used through the model should be created. For models with a large number of assumptions, consider recording these on a separate control sheet. For the highest risk assumptions, a record of authorisation/sign-off must also be kept.

**All models must include a version control sheet:** this holds information on the file name, date, name of analyst making the change, sheet on which change is made, details of the change and record of subsequent QA.

For each major change, save a new version of the file with a new name. Try to use a naming convention that allows files to be sorted in chronological order and always include the version number in the filename (e.g. V1/V2, V1\_1/V1\_2). If a version of the model reflects a particular scenario, include a short description of this scenario in the file name or as a tag to the file.

**Input data to the model:** liaise carefully with those supplying inputs to ensure a full appreciation of scope, definitions, etc. As a minimum, record the following information for input data: who supplied it and when, a link (if available, e.g. if it is published data), its level of quality assurance and if/when it requires updating.

Input data in the spreadsheet should be exactly as supplied by the supplier; modellers should not be making manual adjustments to supplied data and any amendments should be done by auditable formulae within the model, not prior to the data going into the spreadsheet.

**Distinguish model imports from model exports:** the approach here will be different depending on the size of your model (e.g. colour coding of cells within sheet versus separate tabs). Colour coding of cells/tabs can either be agreed within teams/units/divisions, or simply described on an information sheet so that it can be understood. Do be mindful of colleagues with visual impairments.

**Where possible, use a consistent time ruler (e.g. monthly, quarterly) across or down all sheets:** adherence to this rule will do nothing for calcu-



lation speed, but will significantly improve the readability of the model and the likelihood of detecting logical errors.

**Presentation Sheets:** must clearly and effectively present the results of the analysis. Ensure that final outputs and the input data/assumptions behind them do not become decoupled. Consider the audience for these sheets.

If the model or presentational sheets are to be published, then the output must comply with Departmental accessibility standards.

**Avoid splitting a model across multiple workbooks:** under certain circumstances, this will be unavoidable; for example, short lead times can necessitate individual modellers working in parallel on separate elements of the model. If data is passed between workbooks, clearly identify imports and exports. Naming links to external files will ensure your model tracks any changes to the location of data in the external file.

## 2.3 Worksheet design

**Each column should have a single purpose:** particularly in models containing a large number of rows, a reviewer should be able to read the formula in the top cell and assume this is consistent all the way to the bottom (i.e. avoid a mix of different formulas and hardcoded values). Any exceptions to this must be clearly marked. And this does not over-ride the need for quality assurers to check sums at the bottom of columns or end of rows to ensure formulas are working.

**Provide a clear and unique label for every column which is intuitive for a new user and includes the unit:** some calculation steps can be difficult to describe, but investing the time will reduce the risk of making faulty assumptions and assist the reviewer at a later stage. In addition, being consistent in labelling conventions will assist in using global find and replace if column labels require updating.

**Avoid unnecessary columns of data:** if a column of data does not impact on calculations/aid the building of the model, review its inclusion.

**Avoid hiding rows, columns and worksheets:** generally, any data used in calculations should not be hidden. This is permissible during the model development stage, but not if a model is going to be shared with others.

**Calculation order on ‘workings’ sheets should flow left to right and top to bottom:** it is extremely helpful for a model reviewer to be able to assume that ‘reading order equals calculation order’. Any exceptions to this should be minimised and clearly marked.

**Use auto-checks wherever possible:** for example, check that row and column totals are consistent across worksheets. Consider use of a single check sheet

which holds all checks of internal consistency and uses conditional formatting to identify errors.

## 2.4 Formula design

**Keep formulas as simple and efficient as possible:** if possible, break down very complex, multi-line formulas into smaller parts to aid the reviewer (smaller calculation steps may be required if you cannot explain a formula in a short amount of time). However, complex design is sometimes necessary, particularly when increasing model automation, in such cases, also add comments to help somebody else picking up the model.

**If possible, use a limited set of Excel functions in the model:** avoid functions where the logic is difficult to follow.

**If possible, avoid Excel functions that are sensitive to structural changes in the model:** some functions (e.g. VLOOKUP and HLOOKUP) can stop functioning as desired because of the insertion or deletion of a row or column and alternative functions (e.g. INDEX-MATCH) should be considered if the spreadsheet is likely to experience structural changes.

**Minimise use of nested IF statements and array/CSE (CTRL+SHIFT+ENTER) formulas:** nested IF statements are difficult to de-code, reduce a model's transparency and multiple combinations are often not properly tested. Large array/CSE formulas can slow down calculations and can be difficult to understand – use if the steps required to avoid using an array add unnecessary complexity. Single-cell array formulas are less error prone than multi-cell array formulas.

**Avoid embedding constants in formulas:** this can conceal a model assumption and reduce a model's flexibility. Instead, enter the constant in a clearly labelled cell, and link formulas to this cell. Then, if you need to, by changing one cell, the constant is changed across all linked cells. In certain cases, however, embedding constants is permissible where they cannot be changed (e.g. 12 for months in a year).

**Reserve use of the ROUND function for the 'Presentation' sheets:** perform all calculations on unrounded values.

**Avoid using unnecessary parentheses in formulas:** unless they improve formula clarity or are mathematically necessary.

**Do not use current sheet references in formulas:** these lead to cumbersome formulas which are harder to review.

**When switching signs, use '-1 \* a' rather than '-a':** this makes the sign switch as obvious as possible to the model reviewer.

**VBA/Excel macros:** only use if they add significant value to the model or are necessary (e.g. iterative calculations). Any code must be clearly commented and there must be sufficient capability within the team/unit/division for quality assurance.

Annex A contains a spreadsheet standards checklist for modellers to use when designing models.



## Chapter 3

# Good practice for Excel-based modelling

### 3.1 Introduction

This Chapter contains *suggestions* for taking advantage of Excel's strengths in a modelling context. These are separate from the rules set out in Chapter 2 in acknowledgement of the unique preferences of individual modellers for building models in Excel.

We also provide links to good practice spreadsheets within the Department, and links to external guidance on spreadsheet modelling for reference.

### 3.2 Good practice in Excel

**Control sheets:** if sharing a large, complex model, consider including a flow diagram, for example using Excel 'SmartArt', to graphically illustrate the model's structure.

**Do not repeat calculations unnecessarily:** even if it is simple to replicate the calculation, it is generally best practice to link back to the source calculation. During model development, there may be need to do repeat calculations, but this does introduce a verification risk, so care should be taken.

**Minimise inter-linking between sheets:** in particular linking should only be in one direction and not back and forth across worksheets. Excessive inter-linking can be an indication of poor logic in the design of the model and can complicate the review process.

**Presentation sheets:** adding a ‘control’ element to a presentation sheet can enable users to manipulate the dynamic aspects of a model and see the results without having to shuffle sheets. This also serves as a high-level QA check of the model’s functionality (i.e. do outputs change as expected).

**Save files as ‘read-only’:** ensuring Excel files open as read-only enforces version control as a new version must be saved if changes are made.

**Use hyperlinks to share models rather than sending by email:** this prevents there being multiple versions of the same model in existence.

**Display column headers and row labels only once in a freeze pane:** include any internal consistency checks (e.g. TRUE/FALSE) in this freeze pane so they are clearly visible.

**Do not merge cells:** merged cells can affect the selection of columns and/or rows, are problematic for macros and unmerging cells is time-consuming. As an alternative, use the ‘Center Across Selection’ option from the ‘Format Cells’ dialogue box to achieve the same presentation.

**Use Named Ranges:** naming individual cells or ranges makes formulas easier to read and understand (e.g. =PRICE\*VAT is easier to understand than =A1\*B1). Carefully manage named ranges using the Excel Name Manager and ensure they are global (rather than local) and well-documented on information sheets.

**In Excel 2010, use the ‘Format as Table’ option for your data:** in this format, formulas entered in the top cell are automatically copied to the bottom of the column. In addition, any formulas that refer to cells in the table will show the column heading in the formula, which improves formula comprehension/clarity.

**Excel formulas can contain spaces:** these can be used to separate arguments in complex formulas to improve readability. Also the ALT-RETURN combination can be used to start a new line in the formula bar.

**Make use of Excel’s Formula Auditing tool on the ‘Formulas’ tab:** use the Trace Precedents and Trace Dependents commands to graphically display the relationships between cells and formulas with tracer arrows. The Evaluate Formula tool can also be used to debug a formula by evaluating each part of the formula separately.

**Where possible, make use of Excel keyboard shortcuts:** as well as improving the speed at which you can work in Excel, keyboard shortcuts can reduce the risk of making errors (e.g. CTRL+SHIFT+UP/DOWN/LEFT/RIGHT is a reliable way of selecting data ranges).

### **3.3 Good practice spreadsheets**

There is now an area on DfE's intranet dedicated to quality assurance on analytical models, where this document and that on the QA guidance are located. As modellers find examples of best practice, or indeed consider their own models exemplars, these should be shared with others via this page. There is a strong need for modellers to share and learn from best practice, and senior analysts should encourage the use of this area.

### **3.4 Useful links**

- The FAST Standard organisation
- The Spreadsheet Standards Review Board
- List of Excel 2010 keyboard shortcuts