How to Guide – creating CSV files

Purpose

We often have to disclose datasets to comply with our obligations under the Freedom of Information Act 2000 and the Reuse of Public Sector Information Regulations 2015. CSV is commonly accepted to be a secure and reusable format in which to do this, if created in the right way.

A dataset is a collection of factual information in electronic form about the service or functions of a public authority like the ICO.

When we disclose datasets, we fulfil our obligations to provide information in a reusable form by converting and disclosing that information as a CSV file.

For more information please see the published ICO guidance: How to disclose information safely.

Instructions

The key thing to remember is that a CSV file is a basic file format which does not support advanced features such as hidden rows and columns or pivot tables. This is why it is useful in circumstances where you want to make sure metadata or irrelevant information is not included and why it might not be so useful if you want to retain any metadata.

Using Excel as an example, if you hide or filter data and then save the file as CSV the hidden data will be present and visible in the CSV file. It is necessary to delete the data from the original Excel file to ensure that only the correct data is present in the CSV file.

These instructions will help you to achieve that. The best way to check is to open the CSV file and review the contents. Make sure that data you know needs to be removed has been removed.

If you are ever not sure that this process has worked you should raise this with a manager before making a disclosure. It is important that you do not disclose exempt or irrelevant information.

1. Save a working copy of the Excel spreadsheet locally onto your desktop or into an appropriately named folder.

2. Excel spreadsheets may have multiple tabs containing information, for example; casework statistic reports may have a 'raw data' tab, they may also have a tab containing a pivot table containing relevant statistical information.

3. You may need to tidy up the data by deleting unwanted columns or rows, but only if these are definitely not required. **You should use the 'delete' function to do this. Do not simply hide or filter information.** You can isolate information by filtering but you should then delete what you don't want. This is because if you attempt to copy data to a new workbook this can also copy across data you have previously hidden.

4. You might also want to amend data fields or edit the column headings to display and sort the data into a logical format that can be easily interpreted by the recipient.

5. When you are satisfied with the way the data is presented, follow the instructions below for each sheet in the Excel workbook that you will supply as a CSV:

   Copy the data into a new Excel file > then file > save as > save to disk > save to desktop or documents folder > save as type > from the save as type drop down list select 'CSV (comma delimited) (*.csv)' > amend the file name if necessary > finally select save.

6. When you click the 'X' to close the Excel spreadsheet you may get a Microsoft Excel pop-up warning message. Provided you have made no further changes to the data, Click 'don't save' and Excel will close. If you clicked 'save' by mistake you will get another pop-up warning about features which are not compatible with CSV. Click 'Yes' and the file will save any changes to the previously saved CSV file.

7. If the Excel spreadsheet has multiple tabs you will need to save each one individually as a separate CSV file by repeating steps 3 to 6.

8. Now check each CSV file for information you did not mean to include. You can do this opening each saved CSV file in Excel and by clicking the following:

   File > check for issues > inspect document > document inspector > inspect.

   You will see that the check is for a number of different types of information. If it finds, for example, hidden data that you did not mean to include then you should delete the data and repeat steps 3-6.

9. You should check the contents of the CSV file yourself. Do this by opening it and check it contains what you wanted it to.

10. As a final check, you should also ask a peer to double check the data and explanation before your disclose it. Label your file clearly and ask your peer to check it. You should either provide them with a copy of your accompanying explanation or be able to explain to them what it should contain.

    If required, you can import your saved CSV files to CMEH in the same way that you would add any other document to a CMEH case.

11. You should add a note to your case to confirm that this procedure has been followed when creating your file.