How To Publish Information Safely

Removing personal data from information requests and datasets

Overview

There are several ways personal data can be stored in a file and disclosed by mistake.

However, exporting data to simple file formats such as CSV (Comma Separated Value) and using the inspect worksheet function can highlight potential unauthorised disclosures.

Mistakes can be made in the redaction process and other file types such as email or images can contain additional meta-data.

Hidden data – Hiding in plain sight

The simplest case of data being disclosed in error can occur when it is not immediately visible on screen but elsewhere within the file. This can be due to design choices or formatting styles.

Within a template, a user might have chosen to hide certain data by setting the font colour to the same as the background (e.g. white on white or black on black). While this prevents personal data being disclosed on a printed version of the file it will still remain within the source file.

The personal data is at risk of unintended disclosure if the electronic version is distributed. Highlighting the text or changing the font colour will expose it.

This data is at risk if the electronic version is distributed. Highlighting the text or changing the font colour will expose it.

Redaction

When disclosing information under FOI or in response to a subject access request, it may be necessary to redact certain information. In particular you should consider whether the information requested contains personal data. The ICO's advisory note - which places a moratorium on the disclosure of original source spreadsheets to online platforms – includes a number of useful reminders and links on the safe disclosure of information.

Checklist

What you need to consider when disclosing information - a summary of the ICO guidance, for further information please see the What is personal information guidance.

Spreadsheet e.g. xls(x), ods

- Convert the file to CSV
- Use Document Inspector in Excel
- Are you sure you know where all the data is?
- Are there hidden columns or rows?
- Are there hidden work sheets?
- Do pivot tables contain linked data?
- Do charts contain linked data?
- Are there formulas which link to external files?
- Is there meta-data that should be removed?
- Is the file size larger than you might expect for the volume of data being disclosed?

Word processor e.g. doc (x), odt

- Are there comments within the document that should be removed? E.g. Section 40(2) exemptions
- Does the document contain a version history?
- Do pivot tables contain linked data?
- Do charts contain linked data?
- Is there meta-data that should be removed?
- Does the document title or filename contain personal data (e.g. Letter to John Smith)?
- Has a header or footer been automatically added to a printout?

Presentation e.g. ppt(x) odp

- Are there presenter notes that should be removed? Eg Section 40(2) exemptions
- Do pivot tables contain linked data?
- Do charts contain linked data?
- Is there meta-data that should be removed?

PDF

- Are there comments that should be removed?
- Are all redactions effectively applied?
- Is there meta-data that should be removed?

Email e.g. mbox, msg

- Is there data within any attachments that also needs to be redacted?
- Is there meta-data that should be removed?

Image and video e.g. jpg, avi

- Is there attached EXIF data?
- Is there personal data that needs to be obscured (e.g. faces of third-party individuals)?

Security Education and Awareness Centre
Supporting Your Security Education and Awareness Journey

For in-depth guidance, read the ICO's How To Disclose Information Safely