# 04. Optimising files like PDF, PPT, DOC and XLS

## Purpose

This page addresses the problems that can arise from using PDF, DOC, XLS, PPT and similar file formats to communicate on the web.

## Rules

For every file on EU and EC websites, ensure that

- the file has a meaningful name, possibly dash-separated (e. g. *f4538945ertg_bis.pdf* is a bad name, *commission-module-for-environmental-protection-ID-4538945.pdf* is a better name)
- specify a meaningful meta description for the file
- specify an author for the file (either a person or a DG)
- ensure that the file is reachable from a page on the website
- once the file is not relevant any more remove it from the website
  - The only exception to this rule is when the file has to be kept online for legal or historical reasons. In that case an archived version of the website has to be created and the page can be moved there.
- ensure that the links contained in that file are reachable.
  - If they aren't try to correct the original file
  - If you can't edit the original file anymore, set the X-ROBOTS-TAG header of that file to *index, nofollow*. In this way the HTTP response of the file will discourage crawlers from following the links contained in the text
- keep the file size as small as possible
- always specify the file size from the page that links to that file (usually this is done automatically by the CMS)

## Guidelines

As a general rule, this kind of files should be avoided. In fact, they are a nuisance for the user as they:

- require a specific third party software to be viewed (e. g. Adobe Acrobat, Microsoft Power Point and so on) which might not be installed on the user's machine
- might imply a security risk (e g. for PDFs)
- take extra time to be downloaded and interrupt the navigation flow
- are hard to maintain in time (changes require using third party software)
- are hard to be made accessible
- might be cumbersome to parse for screen readers and web spiders

On specific occasions (e. g. in case of web forms or complex tables), such files might nevertheless be the appropriate choice. Please, consider the following points:

- if possible, offer an alternative HTML version of the file
- for PDFs
  - make sure they follow the PDF/UA standard (this request can be added to contracts with third parties)
  - add extra meta tags to describe the page as clearly as possible
  - keep the layout simple
  - don't use images to represent text, as they require an extra step to be parsed
  - Never use xxx for user data's placeholder. Use the a more informative version (e. g. *insert your name* / *surname* / *age instead*)

## Accessibility

When creating and linking these files you must respect the accessibility guidelines.

## Contact and support

If you require further assistance, please contact:

Comm Europa Management

European Commission
DG Communication
Unit B.3
Europa Web Communication