Removing Metadata from Word Documents & PDFs

What Metadata Is and Why It Matters

Metadata is hidden information stored within files — like author names, document history, comments, tracked changes, revision dates, and even location data. It can reveal sensitive details unintentionally and should always be removed before sharing documents externally.

Common Metadata Found in Files Microsoft Word Documents (.docx)

- Author and company names
- Hidden comments and tracked changes
- Document revisions, timestamps, and file path
- Custom properties or template information

PDF Files:

- Author, creation date, and software used
- Hidden text layers and annotations
- Embedded links or tags from original source files
- GPS or image EXIF data (if the PDF includes photos)

How to Remove Metadata in Microsoft Word

Option 1: Use the Document Inspector (Windows desktop version)

1. Open the document in Word.
2. Click File → Info.
3. In the right-hand panel, click Check for Issues → Inspect Document.
4. In the Document Inspector window, check all the boxes (especially Comments, Revisions, Versions, and Annotations).
5. Click Inspect.
6. Review the results, then click Remove All next to each category you want to delete.
7. Click Close, then Save As to create a clean version for sharing.

Option 2: For Word on Mac

1. Open your document.

2. From the Tools menu, select Protect Document.
3. Check the box labeled Remove personal information from this file on save.
4. Save the document. This strips identifying metadata automatically.

Option 3: Manual Cleanup (All Versions)

- Accept or reject all tracked changes.
- Delete all comments (Review → Delete → Delete All Comments in Document).
- Ensure “Author” and “Company” fields in File → Info → Properties → Advanced Properties → Summary are blank.
- Save the cleaned file as a new copy.

How to Remove Metadata from PDFs Using Adobe Acrobat

1. Open the PDF.
2. Go to File → Properties → Description and clear all fields.
3. Then go to Tools → Redact → Remove Hidden Information.
4. Check all boxes (comments, metadata, file attachments, etc.) and click Remove.
5. Save the sanitized version under a new file name.

Removing Metadata from Google Docs (Before Downloading as Word Files)

When downloaded as Word (.docx) files, Google Docs can retain:

- Document owner name and email address
- Revision and comment history
- “Last edited by” details
- Internal sharing permissions

Best Practices Before Downloading Option 1: Create a Clean Copy

1. Click File → Make a copy in Google Docs.
2. In the new copy: o Delete all comments and suggestions (Tools → Review suggested edits → Accept all). o Remove all collaborators (Share → Remove access). o Ensure your name/email are removed from the document title and content.
3. Download as Microsoft Word (.docx).
4. Open the file in Word and run the Document Inspector (see Section 3 above).

Option 2: Export as PDF

1. Click File → Download → PDF Document (.pdf).
2. Most collaborative metadata will be stripped automatically.
3. Use Adobe Acrobat’s “Remove Hidden Information” or BleachBit/ExifTool for a final clean.

For Highly Sensitive Documents

- Avoid personal names or internal links in content or comments.
- Use a neutral, non-personal account (e.g., admin@organization.org) as the file owner.
- Rename files generically before sharing (e.g., Policy_Guidelines_CLEAN.docx).
- After download, recheck metadata in Word before distribution.

Automation for Google Workspace Admins

- Disable “Show editors” in version history for shared links.
- Set default link-sharing to “Restricted” or “Anyone with the link (Viewer).”
- Use Google Workspace DLP rules to flag documents with personal data before export.

Organizational Best Practices

- Always inspect documents before sharing. Make this part of your publication or communication workflow.
- Never rely on “print to PDF” alone. Some metadata can survive that process.
- Store original versions securely on encrypted storage like Tresorit, not in email threads or shared folders.
- Train all staff on why metadata matters as part of regular cybersecurity and privacy awareness.

When in Doubt

If a file contains sensitive content (e.g., internal strategy, donor data, or legal documents), sanitize it twice — once in its native format and again after conversion to PDF. For highly sensitive reports, share only through encrypted channels like Tresorit or Signal.

Revision #3
Created 30 October 2025 20:54:01 by Josh
Updated 30 October 2025 22:05:39 by Josh