

## The Top Ten Hidden Data Threats

These are the top ten hidden data issues according to our research. All issues were ranked according to the frequency of occurrence, the amount of sensitive information that could be exposed, and the overall risk to information security.

### 1. **Ad Hoc Review/Tracked Changes**

Microsoft Outlook automatically enables Tracked Changes without warning when a Word, PowerPoint, or Excel file is attached to an email. As a result, 20-30% of the Microsoft Office files on the Internet are Tracking Changes. This may be even higher in the U. S. Government, where Outlook is the preferred email client.

### 2. **Embedded OLE Objects (esp. spreadsheets)**

You didn't just copy the chart, you copied the entire workbook. Object Linking and Embedding (OLE) is the glue that lets us embed the data from one application (like Excel) into another application (like PowerPoint). Unfortunately, when you embed an OLE object, you embed a complete copy of the original file, not just the visible data.

### 3. **Layered objects**

Although we think of our electronic documents like a printed document, the electronic version has many layers. While working with your data, it is easy to place objects on top of other objects, especially when using copy-and-paste. The object in the background can still be recovered, even though it is not visible.

### 4. **Resized objects**

Consumer grade digital cameras offer resolutions in excess of 7.1 mega pixels. When that picture is imported into our electronic documents, the entire contents of the picture is embedded in the document. Resizing the picture to fit does not remove any information from the file, and the full content of the picture can be recovered. Excessive resolution is a leading cause of oversized files.

### 5. **Cropped objects**

Users frequently crop data they do not want to use in their electronic document, but cropping does not remove any data from the file. Anyone can use the same cropping tool to uncrop the object and reveal the hidden data.

### 6. **Paste and Paste Special issues**

Paste is a very powerful feature that automatically converts data on the fly to meet the requirements of the application. This can result in special constructions that can conceal information that was not intended for the electronic document. Data conversions are most likely to occur when copying information between applications or different views in the

same application. Paste Special allows the user more control over the Paste function, but if the user does not understand how Paste works, they could easily create one of these dangerous data constructions.

#### 7. **Hidden data ported to Adobe PDF documents**

Adobe's Portable Document Format (PDF) has an undeserved reputation as a safe file format because it works through a print driver. PDF documents are not the same as hard copy printout. PDF documents contain metadata, layered images, and other data that is hidden in the original document.

#### 8. **Improper Adobe PDF redaction**

Printing to PDF is NOT the same as printing to paper. If the data is not properly redacted in the original document, then it will not be redacted in the PDF version of that document. NSA and Adobe have published guidelines for redacting Microsoft Word documents, but those guidelines do not address many types of hidden data that will port into the PDF document.

#### 9. **Deleted content in documents**

Although the Microsoft "Fast Save" issue is well known, we still see a lot of documents that have been fast saved. The Fast Save was invented to speed up a save when writing to slow media, like a floppy disk. Rather than writing the entire document to disk, only the changes are written and appended to the end of the file. This is why some files get larger even when you are deleting data. This process is also called fragmentation.

Adobe PDF documents also fragment! When PDF documents are edited, new information is added to the file, but old information is not removed. While the "Save As" function removes Fast Save fragments from Microsoft Office files, the Adobe "Save As" function only removes some of the unused data. Editing PDF documents to remove sensitive data is very dangerous.

#### 10. **Highly formatted data**

Data doesn't have to be hidden to be overlooked. The very purpose of formatting is to call the readers attention to certain data deemed to be more important than other data. This occurs frequently when information is posted to the web. Human nature is to focus on the data to be published, while overlooking other data in the document that may not be appropriate.

### **Metadata didn't even make the list!**

Metadata is only one small part of a much larger hidden data problem. ManTech SRS has discovered well over 100 ways that users can lose information inside an electronic document.

A 20-minute, narrated presentation of the Top Ten Hidden Data Issues is available at <http://www.docdet.com/TopTep.pps>.

A streaming video version is available at <http://streams.docdet.com/TopTen.m3u>