White paper

The data you can’t see that can hurt you

Workshare

Change matters

What's inside?

Your guided tour:

Workshare®
Change matters

Introduction

Most employees go about their working day unaware that the business documents they are sending or sharing outside the control systems of their organization contain hidden and potentially sensitive data. Referred to as *document metadata*, it isn't something that's always visible to the naked eye and it isn't something we generally think about.

Hidden document metadata, however, left in files when they're shared can expose sensitive and confidential information, such as document properties, redacted or deleted text, or notes made during a review process. Any of these different elements can put employees and a company at serious risk.

As a minimum, document metadata leaks can lead to loss of confidence in a firm's capabilities, a loss of clients, or disciplinary action. At worst, serious reputational damage, financial loss and lawsuits can ensue, especially if social security numbers or personally identifiable information is involved.

On the face of it, many may think that this is an issue for their company's security experts – and, to an extent, it is – but as we increasingly use personal devices and applications to share business documents, employees must take responsibility for protecting confidential information, sensitive or personal data, and intellectual property from accidental leaks.

So, what exactly is metadata? And what are the best ways to ensure documents released outside your organization don't have it hidden inside them?

This paper aims to safeguard professionals who work with sensitive information, helping to prevent them from inadvertently sharing metadata. We explain what document metadata is, what forms it can take and how to deal with it.

> **Banks, law firms, defense contractors and government departments were all found to be leaking data.**
>
> *'Shoddy data-stripping exposes firms to hack attacks.'*
> *BBC technology | July 2017*

What is document metadata?

A document's metadata is essentially information about that document, including changes made during its development.

When we create, edit or save a document, behind the scenes a rich set of metadata is automatically added. This can include information about how long the document has been worked on; how the file was created; time and date stamps; who the original author was; when the document was last saved; tools used; and a short summary of the document itself. As the document evolves, track changes can also be added and included in the content.

Originally conceived to make it easier to track and find document data, no one argues that metadata is useful when used properly and with an appropriate level of awareness. However, such rich information "hidden" as a file is stored, modified and shared can reveal more information than we would want to share with other people.

For example, edit and review features in Word, such as reviewer comments, track changes and 'undo' features result in a significant amount of risky metadata being included.

Also, not every document starts as new. Workshare's experience shows that up to 70% of documents are 'recycled', i.e. they start life as a copy of another file. People tend to base a new document on a similar one that already exists and they begin an editing process. This is perfectly acceptable. However, imagine billing a customer for a new report or proposal you were creating specifically for them - you would not want the origins of that document revealed inadvertently.

Forgetting or ignoring that metadata exists when documents are shared means 3rd parties can access privileged, confidential information never intended for them to see. There have been numerous high-profile cases involving leaked metadata and the catastrophic effects it can have on businesses.

Proof point

During contract negotiations, a team starts with a standard document cloned from a previous project to save time. They replace "Company A" with "Company B" throughout; have an internal review; and make edits using track changes enabled so co-workers can see the progression of the document.

During this internal review, a clause the team deems unnecessary is removed. They also remove some statements they had put into the dealings with Company A, which they don't want to offer Company B. Later, they add a clause to address some particular internal concerns about dealing with Company B, and notes are made as comments to explain why.

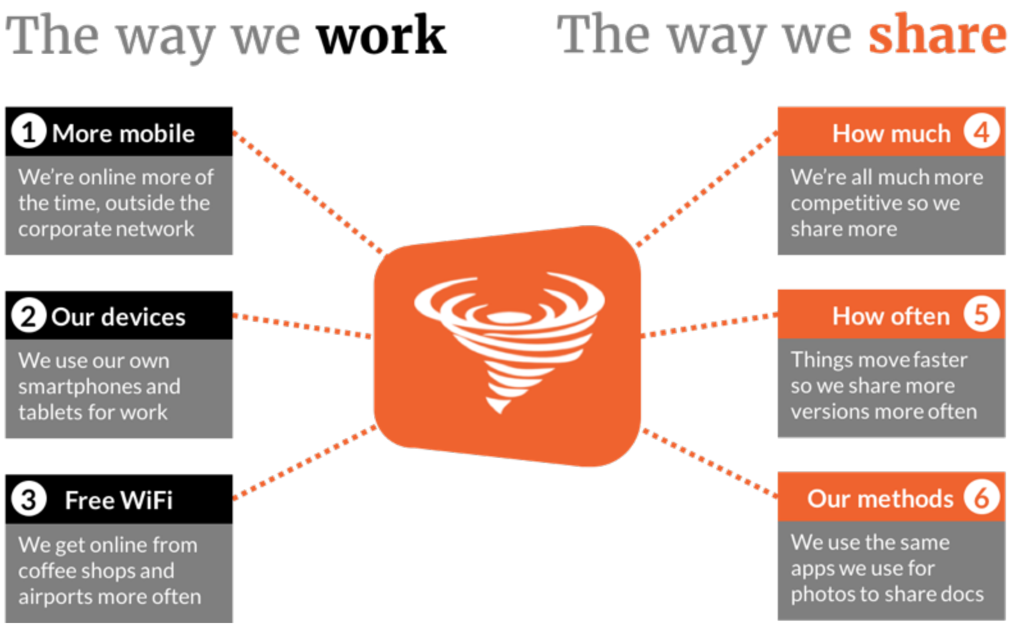
Imagine how damaging it would be if all this information were shared with Company B in the file's metadata.

How the sharing risk emerged

The volume of files being shared electronically between business professionals increases each year. This rise has been driven by adoption of technology in the workplace, of course, but is really a result of increased competition. There are essentially two factors involved in creating a perfect storm in data security terms. Increases in the volume and velocity of file sharing, coupled with cultural changes that have led to a growth in the risk of hidden metadata in files emailed or shared online.

- **The way we work**
  With a seismic shift towards remote or flexible working and an increase in working outside office hours, cultural change has created a higher propensity for risky sharing behavior, without people even knowing they are culpable.

- **The way we share**
  A need for speed with faster document review cycles, wider distribution of information to globally dispersed teams, and greater collaboration among groups of different people contributing to documents.



The way we **work**

The way we **share**

1 More mobile
We're online more of the time, outside the corporate network

2 Our devices
We use our own smartphones and tablets for work

3 Free WiFi
We get online from coffee shops and airports more often

How much 4
We're all much more competitive so we share more

How often 5
Things move faster so we share more versions more often

Our methods 6
We use the same apps we use for photos to share docs

A perfect storm for data security challenges...

Metadata elements and their associated risks

The key to the problem is not that metadata is added to a document, but that it is difficult to fully identify and remove. For example, in Word, adding comments and tracking changes are very helpful to people working on a document. However, when a change is not accepted, it remains within the document even though it appears invisible. These changes can easily be displayed by turning on the "Show Mark-up" view, which can result in damaging situations where external parties see information never intended for them.

Here are six areas in Microsoft Office documents where metadata problems can arise:

1. Review details & change history
2. Document properties
3. Document statistics & file dates
4. Document reviewers
5. Hidden text or macros
6. Custom properties

1. Review details and change history
Applies to: Word, Excel, and PowerPoint documents

Track changes, comments and document revisions, including the last 'undo', help an author understand what additions and deletions have been made by others working on the document. These are tagged with the initials of co-authors. Comments are included to help reviewers make suggestions to the person collating all the final information. Previous versions and fast saves can also fall into this category.

Where are the risks?
Identifiable comments left in documents as metadata can be devastating. For example, a senior member of the team recommending that something be removed or changed before it reaches the intended recipient of the document can survive as metadata if not actively removed. Not only will the recipient see the item that needed to be taken out, they will also see the reviewer wanted it to be removed.

2. Document properties
Applies to: Word, Excel, and PowerPoint documents

Document properties are details about a file that help identify it, including a descriptive title, subject, author, manager, company, category, keywords, comments, hyperlink base, server, network names and anything that reveals a blueprint for a hacker. Document properties display information about a file, so they can be organized and found more easily.

Where are the risks?
The names of authors and organizations in document metadata can divulge sensitive information to 3rd parties. For example, if a document is sent outside your own organization, the author name or company name contained in the built-in properties could be one other than your own. Also, if a document were repurposed or used as a template for a new document, information specific to a previous client (for example, pricing, terms or client names) could be stored as hidden data in the new file.

Metadata elements and their associated risks, continued

3. Document statistics & file dates
Applies to: Word documents only

Document statistics include information on when the document was created, modified, accessed and printed. In addition, document statistics display the name of the person who last saved it, the revision number and the total editing time. Other statistics include number of pages, paragraphs, lines, words and characters.

Where are the risks?
Document statistics can create embarrassing situations. For example, the 'last saved by' metadata shows the last person who edited the document and can create disputes over who actually produced and worked on a document.

4. Document reviewers
Applies to: Word documents only

Document reviewers consists of a list of users that have added or accepted document changes.

Where are the risks?
Document reviewers' metadata exposes who suggested what changes. Removing the names of reviewers can be as important as removing the changes they have suggested. While the reasons for involving someone in a process may be completely legitimate, they may be hard to explain to clients.

5. Hidden text or macros
Applies to: Word, Excel, and PowerPoint documents

Especially in Microsoft Word, metadata issues arise in hidden text, footnotes, white text and small text remaining in files.

Where are the risks?
Either added to documents wilfully or unintentionally, these items are not visible to the eye when reviewing a document but can easily be made visible if they remain in the document.

6. Custom properties
Applies to: Word, Excel, and PowerPoint documents

Custom properties include any property fields added to a document manually or by various programs to help manage and track files.

Where are the risks?
Custom properties are normally specific to an organization. Common types of custom properties are document ID, department, and status. Custom properties can reveal proprietary information or competitive business practices that would be inappropriate to share with 3rd parties, especially clients.

Three simple steps to better file security

In light of these risks, these three key considerations are recommended to take into account before sharing documents.

1. How: Consider the methods used to share documents

The traditional method for sharing company documents has been as email attachments. However, as there have been massive changes in the way people work, especially with the surge of 'Bring Your Own Device', more and more sharing is carried out on phones and tablets, much of which is outside the traditional controls of IT. While IT departments may have policies or technology in place to protect email attachments sent from a desktop, we're now sharing in so many different ways, including consumer-grade file sharing services or mobile devices, these also need to be protected.

IT sanctioned tools and processes must be deployed company-wide, across all devices and platforms. To ensure employees use the tools provided, they must also make working and sharing more efficiently.

2. What: Consider what is being sent when sharing a file

The documents professionals share contain high-value intellectual property, confidential information and even highly sensitive or personally identifiable data. Also, inside the same documents, behind the text or a clickable embedded table, is hidden metadata – track changes in a Word document, notes in a PowerPoint presentation, or confidential financial information in an Excel sheet. Someone could, for example, click through to the complex data and formula held in Excel behind what was meant to be a simple graph in PowerPoint.

By being aware of all the data we may be sharing when sending a file, it's possible to ensure systems are in place to selectively remove the metadata not intended for sharing.

3. Who: Consider snooping and surveillance

The term metadata has entered more widely into mainstream vocabulary. With WikiLeaks and scandals around national security and intelligence agencies making headlines, everyone is more aware of needing to protect their digital footprint.

As the controversy around privacy and snooping continues, protecting clients, high stakes documents and the metadata inside them is each individual's responsibility.

With document metadata hidden inside reports, spreadsheets or presentations shared and stored online, businesses and the professionals in them have the ability and a responsibility to protect the data they are processing.

Conclusion

Document metadata can serve useful purposes for identifying, indexing, and managing documents.

It is critical for us all to understand how metadata is created, where it is stored in our documents, and how it changes, especially when collaboration takes place. All metadata elements can create risk by revealing sensitive or confidential information that may result in discrediting incidents, competitive disadvantage or outright legal action against an organization.

So, what should firms and individuals do about this?

- Be aware of where document metadata exists – understand what it is and how to selectively remove it from documents that need to be shared. The methods professionals use, especially when sending documents or sharing them via a mobile device and networks outside the office, must be sanctioned by IT and fall under company compliance policy. Equally as importantly, these methods must provide tools for metadata cleaning.
- Take control and mitigate the risks around metadata that could be leaked or exposed before it's too late. The first step is to become more conscious of the risks and how to manage them. Metadata can and generally should be considered for removal before distributing a document outside an organization, so check and see if this best practice is being implemented. If it's not, then take steps to ensure your organization and everyone in it has access to metadata removal tools for files shared via email or uploaded to a browser.

About Workshare

Workshare is dedicated to helping professionals compare, protect and share their high stakes documents. Since 1999, Workshare has developed and released intelligent technology for business services firms. Now, more than two million professionals use Workshare around the world.

Since being first to market with the ground-breaking DeltaView technology, Workshare has honed and perfected document comparison software. 15 years of experience has led to the best way to compare two documents. If you would like a demonstration of Workshare's file comparison solution, please call us on +44 (0)20 7426 0000.

© 2014 Workshare. All rights reserved. All product and company names herein may be trademarks of their registered owners.
Sold under a license for U.S. Patent Nos. 7,895,276 and 8,060,575 and 8,977,697.