



What Lies Beneath Your Documents May Embarrass, Hurt or Cost You

White Paper

In the United States alone, over 60 million information workers employ the Internet and e-mail in their daily duties. Most now view e-mail as more important than the phone as a business tool. Yet, as indispensable as these tools have become, they expose most organizations to increasing risks, especially as it relates to the transmission or posting of documents. These documents are often published with hidden business data or confidential text not intended for the recipient or outside world to review.

Rarely does a month go by without the media disclosing that a government entity or business organization had published a document containing content not intended for publication. In many countries, new content and data protection laws prohibit the disclosure or reuse of certain information – The *Data Protection Act* of the UK.

Given that the delivery and publishing of documents via e-mail and the Internet will continue to increase, organizations should protect against the risks inherent in attached or published documents. However, very few organizations have proactive policies, procedures or protections against this type of inadvertent content disclosure.

In summary, what you see is not all you are sharing. This whitepaper details the Top 10 content elements that impart risk, defines paths toward auditing for that risk and suggests rules and procedures which aid in minimizing those risks. Many of these items cannot be detected using existing “metadata scrubbing” products.

Summary: Top 10 Content Elements That Impart Risk

Content Element	Potential Risk
Tracked edits	Exposes changes, contributors to the document
Text hidden from print, view	Reveals internal guidance instructions
Text redacted using font color, graphics	Risks privacy, security compliance
Text in comments	Discloses remarks not intended for 3rd parties
Document profile information	Can harbor privileged categorization data
Passing 1-5 above into resulting PDF	Inadvertently disclose source, contributors, remarks
Previous versions of the document	Reveals entire drafting progression
Last 10 authors, previously deleted text	Names prior 'owners', chronicles all edits to the file
Email attachment, routing details	Reveals sender, subject line used to launch review cycle
Inserting embedded objects	Releases entire editable worksheet, graphic or intellectual property designs

Detailed: Top 10 Content Elements That Impark Risk

Risk #1: Microsoft Word's Track Changes set "not visible" when saving or opening

As document content moves between authoring and reviewing cycles, participants often use Word's Track Changes feature to mark what revisions were made and by whom. Should a document contain substantive changes and/or changes tracked by multiple contributors, these markings often render the document 'illegible' during its final editing phases. To accommodate, Word provides users with a feature to "show" or "hide" the markup while contributors make edits or review results.

If the document's final editor 'hides' the markup, then saves the document, Word stores this view setting (along with several others) as a document preference, thus "hiding" tracked changes from the view of anyone subsequently opening it. The setting is only changed when it is toggled to 'show' changes and the document is resaved.

Each tracked change remains active in the document until it is "Accepted" or "Rejected", thus can be passed along to each successive document created from it. Because the document's tracked changes view setting is 'hidden,' subsequent editors or owners of the document may never know what lies within the file until the initiation of a tracked changes session, or a change is made to "Show" tracked changes. Word 2003 provides a "Security" setting, "on" by default, which forces such changes to be visible when such a document is opened.

Posing an additional challenge is one of version-specificity: Although change tracking has been a feature since Word 2.0, it was not rewritten until Word 2002/Office XP – thus ignoring hundreds of core product functionality enhancements during the process. As a result, "tracked" collaborations between users of *dis-similar* versions of Word consistently challenge the document, initiators and contributors alike. These issues further increase the risk of hiding previously-tracked changes in the depths of the Word binary file.

Risk #2: Text marked with a "Hidden" font attribute or text obscured from view

Setting text as 'hidden' (Format | Font | check "Hidden") is a technique commonly used by document or template authors to include document production or content development guidance in model or shared documents. When such content remains *within* the boundaries of your organization or workgroup, there is little to no risk. However, should this content be distributed *outside* the organization, it can prove embarrassing, perhaps harmful depending on the guidance delivered.

Generally, some type of low-level automation is offered to remove hidden text once the document is finalized, yet either users fail to run these routines or the routines fail to perform completely. For example, complex areas of the document are often missed – text boxes within the header, for example.

Also categorized as 'hidden text' are several formatting foibles that obscure authored text from view. While not hidden in the sense of a font color or attribute, text is nonetheless hiding within non-visible sub-structures:

- Text masked behind table cells where 'Exact' row height is set
- Section settings conceal content in dormant headers and footers (e.g., "Different First Page" or "Different Odd & Even Pages")
- Graphics are formatted "In front of Text" or buried within the drawing canvas thus cloaking text not readily evident

Each of these scenarios occurs when time-consuming and complex formatting tasks cause authors, content developers and administrative staff to take common but dangerous shortcuts: copy or cut and paste from other documents, or opening an old document as the means for jumpstarting the creation of a new work.

Risk #3: Text redacted with a white font color or covered by graphic lines or boxes

When sensitive information must be expunged from a document, it is said to be ‘redacted.’ When redacted from paper, white or black tape is placed upon the content and then pages are copied. Users often extrapolate these *hard copy* methods to the *electronic* document using word processing features intended to deliver print-level functionality: obscuring the sensitive information with a black-filled text box or graphic line, or marking content in a white font color.

While effective to the paper-printed page, this method is not effective when “printing” to a PDF file. In the resultant PDF, the image layer displays the desired effect, while the text layer of the PDF file contains *the original text*. On an unsecured PDF file, text can be easily-retrieved through a Clipboard copy, a PDF-to-Word conversion tool – even turning up in electronic searches. A recent case of this was the release of classified data in a military PDF documenting the killing of an Italian journalist at a military checkpoint.

Risk #4: Text in Comments

Word’s Insert | Comment feature permits a user to affix reviewing commentary to a location in the document. When editing or printing from later versions of Word, these Comments appear as balloons extended from the right-side margin, or can be optionally set to display in a separate reviewing pane. Like tracked changes, viewing comments can be distracting, thus users often suppress off their always-on display limiting the users’ awareness the comments exist in the document and setting the stage for those comments to be inadvertently published.

Risk #5: File | Properties content, such as Title, Author, Keywords, Comments

All Windows applications create files with some viewable array of Document Properties, commonly referred to as “metadata” about the file. In Word, specifically, these properties are automatically completed using the first paragraph’s text as detail. When the document is created using File | New in Word, or when disclosure is intentional, there is no cause for concern. However, when the document is created through reuse of a previous document or action (File | Open, File | Save As ...), the previous (or source) document’s metadata can inadvertently provide more insight to the document’s origin or previous collaborators than is intended or desired.

In most Enterprises, these properties can be used to tag documents for integration within a larger taxonomy or document management system, usually for purposes of classification and easy retrieval via search. If the use of Word’s Keywords property is used to organize documents

as “Adversary” or “Prospective Client,” this disclosure may reveal a posture thought to remain proprietary *inside* your organization.

Risk #6: Passing content elements 1-5 (above) into a resulting PDF rendition

Many organizations and individual document authors have recognized both the challenges and the perils of emailing or delivering editable Office source files. These concerns include:

- Do recipients use the same version of Office applications used to create the source?
- Can recipients open all source files included? For example, not all Office installations include PowerPoint, or Excel.
- Is the owner willing to hand over well-crafted content or branded document designs, giving recipients easy access to *reuse* for their own future purposes?

As a result, Adobe’s PDF file format has become a frequent method for electronic distribution. In recent years, Adobe’s extended integration within Microsoft Office applications – most notably Word – permits the seamless creation and attachment of a PDF file to an email, all occurring within a single-click.

Issues arise, however, when these 1-click operations are inadvertently configured to incorporate *active* functionality from Word:

- Word Comments become PDF annotations, viewable from Adobe Reader
- Document profile information is passed to Properties in the PDF result
- “White” or text obscured with graphic lines and boxes becomes exposable text

Additionally, Adobe’s PDF format can be used to ‘package’ electronic source, permitting the publisher to *include* copies of the source file as attachments to the PDF. This results in the recipient being able to open and reuse the editable source when perhaps that was not the intention.

In situations where the organization’s IT group has eliminated the ability to directly integrate functionality with Acrobat, users simply File | Print to the Adobe PDF, or other PDF print driver. In this scenario, inadvertent use of Word’s ‘sticky’ Print feature – a “Document Showing Markup,” which exposes all comments and tracked changes, or printing “Hidden Text” – delivers all suppressed content into the PDF. Couple these risk factors with the common practice of emailing PDFs *straight* from the authoring application *without* first reviewing it, these content disclosures are no longer protected from view.

Other emerging workflows, such as the enablement of PDFs as a vehicle for passing comments or edits using Acrobat and Reader 7 – do minimize the risk by establishing a change management process that conceals the source, yet these remarks are imparted into the PDF

itself. This means they remain accessible to anyone viewing or printing the PDF unless proper security or distribution restrictions are *also* applied.

Risk #7: Internal Word versions – whole editions of a document, saved within a single .doc file

While most enterprises employ a document management system to manage the versioning progression of Word documents, most are unaware the application *natively* provides just such a feature. All Word editions since Word 97 deliver this capability “out-of-the-box” - you’ll find it at the File menu. “Versions” permit a user to save whole-document editions into a single .doc file container. What’s more, Word can ensure Versions be created *automatically* by enabling the “Automatically save a version at close” checkbox. Once checked, as the document is closed, a new Version is created and stored within the .doc file, no warning given. To the unknowing user, the only telltale sign is that the document’s size expands rapidly – we’ve encountered four-page memos which grew to be 3.5MB.

Once again, these settings become preferences that travel with the document – along with all its previous versions and any commenting, tracking or editing history intact – that is, until all versions are deleted (one-by-one), the checkbox is cleared, and the document is resaved.

Risk #8: Automatically saved data such as “last 10 authors” or previously-deleted text

When saving a Word document, most environments unknowingly capture the last 10 authors and preserve previously deleted text. These are automatic features of Word, and are contingent both upon the version of Word being used and the configuration care taken when the application is implemented. Additionally, the last ten file names and paths under which the document was saved are also maintained.

Exposing previously-deleted text can be controlled by disallowing “Fast Saves”, however, many IT organizations remain completely unaware of the danger this option exposes.

It should be noted that the “Last 10 Authors” listing is an automatic feature of Word and cannot be turned off. Shedding the detail requires the file be converted, the content be moved to a new blank document shell where saving stores only *one* previous author, *or* the binary file be externally edited. It should be noted, however, that Word 2003 no longer stores this detail with the file.

Risk #9: Custom document properties, particularly email routing information

Because Office applications have evolved as personal productivity tools rather than Enterprise or Business-to-Business publishing tools, ease-of-use additions such as the ability to “Send for Review” or “Routing” to multiple participants aid reviewing cycles.

One such workflow automatically locates the originating document that launched the review cycle, then automates merging participants' comments and changes as their annotated file is opened.

These exchanges are made possible by electronic Routing Slips, Custom Document Properties and document preferences that are *automatically* imparted into documents involved in the reviewing cycle. Custom Document Properties (File | Document Properties | Custom) will identify the sender, sender's email address and the subject line of the email – a location where participants in a review are less likely to monitor propriety: they *assume* only internal participants will see it.

Risk #10: Inserting embedded objects, such as Excel spreadsheets, Visio drawings; or inserting 'cropped' pictures

Once again, each application's ease of accessibility and use can expose far more than an organization intends. For example, consider embedding an Excel spreadsheet into a Word document. Generally, Excel is used to 'package' multiple worksheets into a single file, keeping current financial or other data accessible for multiple clients or transactions in progress. While the insertion of an OLE object into a Word document shows only a portion of the spreadsheet, the *entire spreadsheet* is *actually* inserted into the document. Subsequent distributions of this source can prove – at best – embarrassing; at worst, compromising or costly. Of similar concern are Visio drawings that often contain highly-confidential organizational IP such as organization charts, workflow or process diagrams, or engineering plans.

In the case of drawings or pictures, users often leverage the Office applications' "Picture Toolbar" feature called 'Crop.' This tool allows the user the ability to pan into the portion of the picture they want represented. But what users may not realize is that the *whole picture* becomes embedded in the document, with its other elements available to anyone who double-clicks on the drawing.

Solutions

Organizations should adopt a multi-pronged approach to this problem:

1. Conduct a practical risk assessment
2. Raise awareness through education, internal marketing of risks
3. Develop policies and procedures
4. Unite technology initiatives to aid and monitor compliance

Step One: Conduct a Practical Risk Assessment

Undertaking a document risk assessment is an onerous task without goals and an automated process. Start by associating levels of risk to your organization for each of the items named in the Top 10 List. This can focus your audit around areas deemed of most immediate concern, thus prioritizing next steps according to those of highest business need.

Next, identify the documents to be audited. Most organizations have well-controlled document or file management systems, and can identify work product by client, matter or other business relationship. This facilitates a measurable, definable scope for your audit. Additionally, work with your internal practice or workgroups to identify high-profile, high-impact projects where documents must “leave the building,” and documents that are *reused* to create ones that will.

Finally, the sheer volume of documents and discrepancies between them rules a manual approach to your audit both impractical and impossible. Thus, secure an automated process such as Microsystems DocXamine to look across work product types – Word and PDF files being the most widely deployed – making sure it can thoroughly research all sub-structures of the document, can configure and sort those items deemed of most risk to the organization, and can extract and track documents ‘found’ from your various document management systems, as well as detailed reporting.

Step Two: Raise awareness, educate document distributors, propose Best Practices or provide elevated levels of support

In advance of formal policies or introduction of digital rights management systems which control the dissemination of electronic work product, educate your document distributors. Deliver a mandatory 30-minute training session that demonstrates risks found in *active work product*, surfaced during the Risk Audit completed in Step One. This is only effective when the illustrations are *meaningful* and when supporting, referenceable documentation is provided.

This whitepaper, along with other content available from Microsystems, Microsoft and Adobe can form the basis for such documentation. Microsystems DocXtools running on the user’s desktop can detect presence of such risks, and deliver this guidance in the open document.

Establish a hotline, Intranet location or other internal communications vehicle to focus on identifying and mitigating risks found. Arm this response team with expertise on all versions of Adobe Acrobat and Microsoft Word, along with knowledge on dissemination methods such as extranets, Outlook or Web pages, taking care to sensitize them to business issues, risks and practices of your organization.

Step Three: Develop Policies & Procedures

Each of the items on the Top 10 list has both an *intentional* and an *unintentional* purpose. You must develop policies, procedures and guidelines that clarify the collaboration context –

regulatory agency, adversarial, internal – then provide concise yet detailed descriptions on *when and when not* to leverage which functionality.

Step Four: Support with Technology Solutions

Automate tedious electronic review, wherever possible: Microsystems DocXamine and DocXtools can be configured to support your policies and procedures or align with various distribution contexts that may occur from members of your organization.

The Microsystems Advantage

Since 1995, Microsystems has helped more than 650 law firms, pharmaceutical and biotech companies, corporations, government agencies and other document-intensive organizations in North America and Europe achieve efficiencies in producing high-quality documents.

Microsystems specializes in developing state-of-the-art software and providing exceptional education, training, support and customer service. These are essential ingredients for world-class document production both now and in the future.

Knowledge Partnership

Knowledge Partnership is a collection of document clean-up and styling software, high-end document support, education and consulting services. This combination is proven to assist firms in proactively addressing problems associated with Word and PDF documents.

With Knowledge Partnership, productivity gains average 75 percent with 10:1 annual returns.

Software

Microsystems' software analyzes documents to provide comprehensive quality control and automated solutions that address corruption issues and ensure compliance with your firm's best practices.

DocXtools—The production, diagnostic and clean-up tool: DocXtools assesses, cleans, and formats documents, aiding in the production of a high-quality document. Many of DocXtools features are available as “one-click fixes” and aid firms in:

- Faster Document Quality Control
- Automating the building and re-build of Cross-References
- Document Clean-up Functionality
- Improved Document Styling
- Enhanced and Automated SEC Clean-up Functionality

DocXamine– The analysis and quality control tool: DocXamine delivers an automated analysis of Word and PDF documents. The analysis detects common problems, summarizes results and provides recommendations for solution through the use of DocXtools.

Pre-Hire Skills Assessment– The Pre-Hire Skills Assessment module enables the automated evaluation of the advanced Word skills of potential new hires. It is targeted at an individual test taker, lists what he/she did right and wrong and provides a total score. It enables firms to make hiring decisions based upon concrete information and it fosters the creation of targeted training plans for the potential new hire.

Support – Solutions Center, Document Emergency Room

The Microsystems Solution Center consists of a team of document experts trained to support the implementation of Microsystems software and the diagnosis and repair of documents. The Solutions Center also manages the “Document Emergency Room,” which fixes problem documents when fast repair is critical. In most cases, a problem document is turned around in 2 hours or less.

Training & Education

Onsite Training: Microsystems provides client training courses, which are role-based, to build specific skill sets for the effective use of Microsoft Word and Microsystems software. Training is offered at client sites or at the Microsystems’ training facility, which is located in suburban Chicago.

Virtual Training: Microsystems provides convenient training and educational webinars that feature topical information about Microsystems tools, industry-specific document issues, Microsoft Word and more.

On-Demand Training: Learn more about Microsystems tools and Microsoft Word from our extensive collection of on-demand training resources, which include How-to-videos, Word resources and tips and tricks.

Client & Consulting Services

Microsystems’ Client Services team provides more than 100 years of combined experience in the areas of document creation, production and quality control. Consulting services can include precedent document clean-up projects, skills assessment customizations, desktop upgrade planning and educational sessions for Microsystems products.

To begin the first step toward improving the document creation, quality control and delivery processes to meet firm goals, please call 630.598.1100 or e-mail our Sales Team at sales.inquiry@microsystems.com