Top 5 Steps to Document Integrity

Workshare™ - The Leaders in Secure Content Compliance

EXECUTIVE SUMMARY

Companies, government agencies and other organizations invest huge resources developing security policies and procuring protective technologies that point outwards at hackers, spyware and viruses. However, organizations are just beginning to realize that there is another aspect to content security, the inside-out leakage of information. Not only do organizations need to worry about the release of valuable intellectual property, but they must now also deal with increased regulation and oversight on issues ranging from consumer privacy to financial disclosure. All of this in an atmosphere of government and consumer mistrust of business that have raised the bar on both actual intended behavior and the due course exhibited in protecting both consumer and shareholder information.

The scope of this internal or "Inside-Out" threat is staggering. According to recent data, the 200 million business users of Microsoft Office, send over 100 million documents over email daily. This amounts to over 125 documents per user/year*. And this is only taking into account the information share over email, let alone other electronic means. Documents are the lingua-franca of business, and they are constantly moving between organizations (see figure 1). The inside-out threat poses serious risks that have the capacity to cost companies huge sums in law suits, regulatory penalties, lost business, intellectual property infringement as well as unquantifiable damage to that most valuable of assets - reputation. Therefore, the key challenge for compliance, privacy, security and legal executives, is effectively understanding and managing this risk, without disrupting the critical flow of these documents that the business depends on.
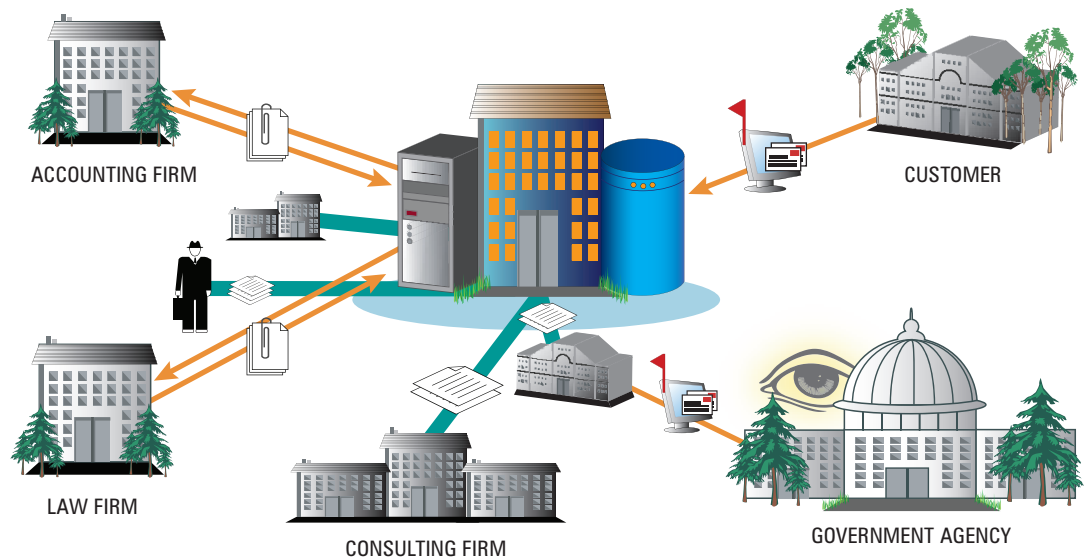

Figure 1: The business document ecosystem and the associated risks

Businesses must manage the risk of 4 types of information leaks:

- Visible information contained in documents and messages
- Hidden information in documents and messages
- Entire documents that must be restricted
- Information artifacts resulting from format transformations

Examples of all these types of information leaks are abundant in the media, and have resulted in international political crisis, regulatory penalties, shareholder lawsuits, lost business and reputation. A summary of recent examples is included in Appendix 1.

Managing the risks associated with the exchange of business documents requires a combination of policy and enforcement. Workshare has developed a systematic approach to help organizations through the process of developing policy and implementing enforcement.

WORKSHARE'S TOP 5 STEPS DOCUMENT INTEGRITY

In today's global business environment, document integrity and security is an ongoing challenge that requires action, measurement and periodic re-evaluation. Only through commitment and focus can organizations hope to manage the risk associated with business documents and their integrity. Workshare has assembled a set of best practices for managing this challenge. These best practices can be summarized as following:

STEP 1: EDUCATION - Understand the three key areas of risk, security, compliance, and accuracy.
STEP 2: ASSESSMENT - Evaluate the level of risk in the organization associated with key business processes
STEP 3: POLICY DEVELOPMENT - Develop ways to classify risk and appropriate mitigation strategies and policies.
STEP 4: POLICY IMPLEMENTATION - Implement the education, systems, technologies and process changes necessary to enforce the policy defined in Step 3.
STEP 5: COMPLIANCE AUDITING - Put in place ongoing and regular auditing of compliance levels and gaps between actual and targeted results.

1 EDUCATE
Understand the threats

2 ASSESS
Level of threat faced

3 DEVELOP
Risk mitigation policies

4 IMPLEMENT
Changes to enforce policy

5 AUDIT
Risk mitigation safeguards

STEP 1: EDUCATION
Understand the 3 key areas of risk; security, compliance, and accuracy.

In order to accurately assess their exposure, organizations must first understand the types of risk associated with the exchange of business documents. Workshare has identified 3 critical areas of risk; security, compliance, and accuracy.

Security - The risk that inappropriate information accidently or maliciously leaves the organization
The four types of information leaks are:

1) **Information that is visible** and violates privacy, intellectual property and/or financial disclosure policies. These policies may be imposed in response to either regulatory or corporate governance requirements. One critical component is understanding international laws and regulations that now govern the exchange of business documents. A non-exhaustive list of US State, Federal, International and Regional statues and regulations is provided in Appendix 2. One key piece of legislation in the United States is Ca SB 1386 which lists six types of information that must be protected, such as a person's first name (or initial) and last name in combination with a unique personal identifier such as a driver's license or social security number.

2) **Information that is hidden** and violates privacy, intellectual property and financial disclosure policies, or can threaten network and systems security. This information can be hidden in track changes, comments, technical data such as network filepaths and links into internal networks. According to research*, three-quarters of business documents contain confidential or sensitive information in the form of this hidden "metadata". Appendix 3 provides a list of the major types of hidden metadata in both Office and PDF file formats.

*(Source: Vanson Bourne, The Risk of Sharing)

3) **Entire documents** that should be restricted. Release of content such as business plans, pre-announcement draft filings, confidential customer records and source code should be restricted. Though often stored in secured datastores, this information is often shared as embedded information in emails and attached business documents.

4) **Format translation artifacts** that are invisible to users, but are still present when converted back into their original format. Examples include white-text, redacted text, non-translated fonts and others.

PDF's retain some of the metadata from the Word file. You can view title, author, document summary, keywords, file location, and comments and tracked changes if they were contained in the original Word document when it was converted. If you don't remove this data before creating a PDF you may share it with the world. PDF functionality is also often misunderstood by the user. A common occurrence is blacklining PDF files – the act of covering up text in a PDF with a black graphic. But when a user blacklines a PDF they often don't understand the content still exists in the PDF and can be viewed by cutting/pasting the Adobe PDF file into Microsoft Notepad.

Compliance - The risk that document processes and exchanges are not adequately defined, controlled and/or auditable

Compliance Regulatory and Corporate Policy – Documents are critical to every business process, including financial filings, and customer and supplier contracts. For every common business process there are associated regulation and internal governance policies that organizations must follow. Compliance requirements fall into a variety of categories such as below:

- Financial reporting regulation such as Sarbanes Oxley; Investment Funds, Companies and Miscellaneous Provisions Act 2005; IASB IFRS
- Privacy and information disclosure regulation as discussed previously and shown in Appendix 2
- Industry specific regulations such as HIPAA in healthcare and BASEL II in financial services
- Protection of intellectual property
- Customer contracts and other documents which impact revenues

It is also important to understand that these compliance requirements can stretch across many documents and business processes. For example, it can be argued that ANY document that impacts the reporting of revenues or costs, must be fully auditable under laws such as Section 404 of Sarbanes Oxley and similar legislation in the European Community, Japan and Australia. This audit requirement stretches to understanding the history of documents throughout their active review process.

Failure to adequately understand the implication of internal and external compliance requirements and how they impact the production and sharing of documents can have serious business consequences.

Tightening regulatory regime

Corporate scandals such as Enron and Worldcom have tightened the noose of global regulation. Legislation such as the Sarbanes-Oxley Act 2002, Basel II and HIPAA demand a higher level of corporate transparency and put content and/or the entire history of documents under scrutiny. This puts more pressure on organizations to think about their vulnerabilities and the risk management processes they have in place.

For a comprehensive listing of relevant legislation see Appendix 2.

Accuracy - The risk that documents leave the organization with incorrect information.

Business documents such as contract, sales proposals and filings go through many iterations and changes and often touch tens and tens of people as they move through their lifecycle and the business document ecosystem. Eric Levinson, systems analyst at international law firm Mayer, Brown, Rowe & Maw LLP, estimates that as many as 50 versions of a single document may be saved to its document management system. This high velocity high volume movement of documents can create inaccuracies for many reasons including:

- Reviewer's comments are simply lost in the shuffle of email exchanges and reviews – In fact, research shows that over up to 75% of documents do not reflect all reviewers' input*.
- Multiple versions of documents and their broad distribution often creates lost and multiple masters, which causes confusion.
- Document management systems simply can not manage and control the flow of documents over email and portals.
- Documents may contain old information from previous versions - over 90% of documents* start life as something else. These old documents often contain old information, that, if included in the final version, can be inaccurate and inappropriate.

No systems or processes can ensure 100% accuracy of documents. However, by understanding some of the common causes of inaccuracies, organizations can then move ahead and manage this risk.

STEP 2: ASSESSMENT
Evaluate the level of risk in the organization associated with key business processes

In this phase of the process an assessment must be performed. This assessment should at a minimum evaluate the risk as defined in step one, the existing policies and processes to manage these risks, or the lack thereof, and user awareness of the risks described.

Security Risk: First, a company can use tools such as those found at www.metadatarisk.org to assess both the information risk across all of its documents. This information should then be evaluated against user awareness to provide a gap analysis of risk versus awareness.

Organizations need to evaluate issues such as:

- How are documents are sent between authors and third parties?
- Who has access to sensitive information?
- What is the level of awareness of risks associated with business documents?
- Is there an ability to restrict documents from external distribution when necessary?
- What visible information is included in what documents?
- How aware are users of these risks?

Compliance Risk: The organization should assess the specific regulations and audit policies that affect each of its critical document types and processes. Next, the policy, process and data available around critical documents should be evaluated to understand the gap between compliance requirements and effectiveness of existing policy and processes. Question the following:

- Do you have a document compliance policy?
- Which regulations affect which documents and what processes?
- Who in your organization is responsible for writing and reviewing – where does accountability lie?
- What policies, internal and external, are involved in the document lifecycle?
- How have your current policies been implemented and verified?
- Can you prove what has been done and why?
- Is there any process in place to provide audit history of documents which fall under regulatory compliance requirements?

Accuracy Risk: Finally, the organization should evaluate the processes it has in place to ensure the accuracy and integrity of key business documents. Organizations should evaluate both the processes and technologies in place to ensure that final documents include all critical user input, and that document masters are maintained and managed effectively. Question the following:

- How precise is the information within a document?
- What processes could compromise the document accuracy?
- Can the content and/or format be altered during the document lifecycle?
- How do users ensure that the master document is not compromised when the document is shared for review?

STEP 3: Develop Risk Mitigation Policies Based on Document Integrity Classifications

Many organizations have developed and implemented content risk classifications. Typically, they are structured something like this:

- ***Highly Confidential:*** Information where unauthorized disclosure will cause a company severe financial, legal or reputation damage. Examples: Financial transactions, customer contracts, business and negotiation strategies, consumer privacy information, intellectual property such as trade secrets.

- ***Confidential:*** Information where unauthorized disclosure may cause a company financial, legal, or reputation damage. Examples: employee personnel and payroll files, intellectual property such as customer and distributor lists

- ***Internal Use Only:*** Information that, because of its personal, technical, or business sensitivity is restricted for use within the company and its close advisors.

- ***Unrestricted:*** Information that in general can be shared, but must still be monitored and managed for risk.

However, in addition, it is recommended that documents be classified not only by risk levels as above, but also by risk drivers. This provides a second dimension of classification. Typical risk drivers are:

- Privacy
- Financial Disclosure
- Intellectual Property
- Industry Specific Regulations

This second dimension allows different handling based on both classifications. For example, a highly confidential financial filing may be fine to email to auditors and financial staff, but not to a credit card processor.

Organizations must also understand how and when documents are classified as above, who is responsible for the classification, and how the classifications will be managed and enforced. Even more important, are the policies for handling each type of document. Examples of this type of policy are the concept of sender privilege and recipient trust, document format policies, hidden data policies and compliance policies.

Sender Privilege and Recipient Trust
In addition, for each type of information, it must be determined both who has the business need to distribute the information and who has a need and is trusted enough to receive this information. For example, the CFO should have ability to share highly confidential information with auditors, board members and members of his team, while his team members may only be authorized to receive this information, but not redistribute it.

Document Format Policies
Depending on the risk classification and driver, documents may be restricted from being shared in certain formats or states. For example, documents with a privacy risk driver may ALWAYS be required to be encrypted when leaving the organization. This is a direct requirement of many approaches to mitigating risk around privacy disclosure laws. Another policy may require ALL final contracts leaving to customers and prospects to be converted to PDF and edit restricted.

Hidden Data Policies

One of the biggest risks to any organization is revealing sensitive, embarrassing or confidential information inadvertently. This data can all too easily be made available to competitors, the press or clients when documents are sent beyond the company perimeter. Policies should be implemented that mandate all staff to remove any hidden data from all documents before emailing to certain colleagues or outside of the organization.

Audit Policies

Once an organization is aware of relevant compliance legislation, it should also put in place policies that enable a visible audit trial of the lifecycle of important documents to be created. This trail should state when the document was created, how and by whom. It should also reveal the changes that were made to it, by whom and at what time. This audit trail must be easily accessed and understood, so that the organization can respond promptly and confidently to any audit, legal, or regulatory request to produce this information.

Once a set of classifications, drivers and policies have been formulated, the organization is then ready to begin the process of implementing and enforcing these policies.

STEP 4: POLICY IMPLEMENTATION

Implement the education, systems, process and techical changes necessary to enforce the policy defined in Step 3.

Compliance officers and security staff must now find ways to ensure that policy is adhered to. This involves implementing a number of changes across the organization:

1. Educational changes

There are two audiences that need to be addressed. Senior management – who need to buy into the overall initiative, understand why policy enforcement is critical, communicate this to all staff, and lead by example. Other staff must learn and understand why they need to follow the new policies being introduced, and see these as an opportunity for the company to be more secure, rather than a threat to their current way of working.

2. Process changes

New policy guidelines also involve a change in the way the organization actually functions (i.e. the processes it follows to carry out tasks). For example, management may stipulate that all bid proposals made by a company must be sent to an overall owner who is responsible for reviewing all sections of the bid, ensuring that its contents are accurate, and that sensitive data is not contained within it, (such as details of an previous bid, comments about the prospective client, or internal discussions regarding the overall proposal strategy). Process changes help enforce policy directives by ensuring the very way in an organization carries out its day-to-day business supports its overall document integrity objectives.

3. Technical changes

It is not enough to rely on staff to follow new guidelines. Wherever possible, the introduction of automated solutions must play a critical part in ensuring that the newly defined policies are followed. These technologies should be easy-to-use, customizable according to the level of user, and where possible feature end-point and network level implementation and some form of centralized policy management and control. They should allow management to restrict the use and distribution of documents, lock-down sensitive content, and automatically remove hidden information according to the policies developed. A technology-centric approach can help ensure that policy transgressions are kept to a minimum. Whilst technology can act as an effective enforcer of process changes, it can also help to educate users as to the risks they are exposed to as they create and distribute documents inside and outside the organization.

STEP 5: COMPLIANCE AUDITING
Put in place ongoing and regular auditing of compliance levels and gaps between actual and targeted results.

Organizations must put in place mechanisms to both monitor and audit the enforcement, appropriateness and effectiveness of their document integrity safeguards. These include:

▣ Periodic review of recent and forthcoming regulations rules and legislation which may change risk profiles and policies. This will ensure the organization does not become complacent and stays ahead of changes in the regulatory environment.

▣ Regular audit of the use of information security technology – how are employees and others accepting and using technology on a day-to-day basis.

▣ Ongoing reviews of the classification of documents and their respective users in order to ensure that policies are keeping up with organizational changes. For example, a reorganization at board level could mean that certain members have enhanced access to sensitive documents whilst others could find their access reduced. Equally, a management consultancy could be granted specific access to certain content as part of their strategic review of the company.

▣ Regular audits of compliance levels across the three critical areas of risk: security, accuracy and compliance. This could involve reviewing "sample" sets of documents or emails at random, or more empirical analysis to track how many Office documents left the company perimeter containing hidden data or a visible content violation over a certain period.

CONCLUSION
The "five step" approach is not intended to be an all-encompassing answer to document integrity concerns, but rather a series of best practices, highlighting the key areas to consider: understanding the areas of risk, assessing these within the organization, developing policies as a result and implementing them, and finally carrying out regular audits to ensure these are being followed.

Document integrity is an ongoing issue that requires action, measurement, and periodic re-evaluation. Only through commitment and focus can organizations manage the risk associated with business documents and their integrity.

APPENDIX 1: Examples of recent information leaks

▢ **November 2005: Australian bank email forces internal review**
Two analysts contacted Westpac to alert executives of an embarrassing bungle that forced the bank to bring forward its results announcement and halt trading of its shares. It mistakenly sent an e-mail containing a template which revealed financial results to 37 analysts in 16 broking firms, before finalization and lodgement with the Australian Stock Exchange (ASX). The template contained adjustments for the previous period, and despite blacked-out squares, some analysts easily uncovered these. The "hidden data" revealed a breakdown of the 2004/05 financial year's earnings, and operating margins. The bank was forced to undertake an internal review and both the ASX and the Australian Securities and Investments Commission were advised of the situation. As a result of this unintended divulging of information, the bank started to hear some speculation in the market that information on our results may be circulating. Westpac had no option but to bring forward the announcement of its financial results as a precautionary measure because there was a possibility that some details of the result may have already been in the market.

▢ **October 2005: United Nations Report causes further turmoil in Middle East**
The complicated political landscape in the Middle East took a new twist when after a 53-page report was sent to the Security Council by a UN investigator. The document accused Syrian officials and their Lebanese allies of assassinating the Lebanese Prime Minster, Rafik al-Hariri and 20 others in a February 2005 truck bombing in Beirut.

At the last minute, UN officials deleted key information contained in the accusation: one witness' testimony that Syrian President Bashar Assad's brother and brother-in-law were among the main plotters of Hariri's killing. Instead of naming names, the final version of the report referred to the alleged plotters in general terms - as "senior Lebanese and Syrian officials." However, an earlier version stated that a witness had identified said officials, including Maher Assad, the president's younger brother, and Assef Shawkat, the President's brother-in-law and head of Syrian military intelligence.

The deleted names came to light after reporters checked previous versions of the Microsoft Word document that contained UN investigator Detlev Mehlis' report, which was e-mailed to the media. The document caused shock around the world with US and UK officials demanding an investigation into these accusations.

▢ **September 2005: UK Government blunder again over terror law**
The UK government was once more in trouble over dodgy document management, with an apparent split within the government over new hard-line anti-terror laws exposed by a letter from Home Secretary Charles Clarke.

The letter was sent via email as a Word document to members of the opposition, and appeared to support controversial plans to hold terror suspects for up to three months without trial. Anyone applying the Microsoft 'track changes' function was able to see Charles Clarke's original wording which expressed concerns over some of the measures.

A paragraph which was deleted from the final version of Clarke's letter reportedly read: "The case for some extension is clear, though I believe there is room for debate as to whether we should go as far as three months. I'm still in discussion with the police on this point."

▢ **May 2005: US military security gaffe**
A serious breach of US military security occurred when classified information was revealed by a simple copy and paste of a document from a PDF format.

The document was a report written after an investigation into the death of Italian citizen Nicola Calipari at a checkpoint in Iraq. The document contains both classified and unclassified information at the incident, which took place in Baghdad on 4 March 2005. The US military later removed the offending document from the Internet, but not before it had been copied and republished on several Web sites.

The military made an error when it chose to simply black-out certain words and paragraphs from the original classified document instead of removing the actual information. This meant that if the document was read or printed, the 'censored' information would be safe. However, by selecting the document text and using the copy and paste function, the document could easily be reproduced in its entirety on any word processing application.

Workshare™

APPENDIX 2: Regulatory Compliance Matrix

Companies must comply with an increasingly complex web of rules and legislation. These regulations target specific threats and industries but a coherent overarching standard for companies to adhere to is still absent.

Worryingly, many companies, according to a recent CIO/PWSC study of IT professionals, believe regulations are "toothless". This could not be further from the truth. Stringent financial penalties and/or imprisonment are a possibility for violations of regulations such as Sarbanes Oxley or HIPPA. For example:

▫ **Health Insurance Portability and Accountability Act (HIPAA) 1996.**

 Example Breaches
 ▫ Misuse of individual identifiable health info
 ▫ Failure to comply with security requirements

 Penalties
  Failure to Comply
  ▫ $100 per violation
  ▫ $25,000 maximum for all violations of a single requirement
  Wrongful Disclosure
  ▫ $50,000 and/or imprisonment for up to 1 year
  ▫ $100,000 and/or imprisonment for up to 5 years if under false pretenses
  ▫ 250,000 and/or imprisonment for up to 10 years if intent to sell information

Below is an overview of key regulations relating to the security and/or integrity of information that companies must comply with covering the North America, Europe and Asia. Please note that this is not a comprehensive global listing.

ALL INDUSTRIES

▫ **APEC Online Privacy Protection Framework** - Currently being developed with the objective of facilitating electronic commerce between the APAC countries and the rest of the world without compromising consumers' personal data.

▫ **Basel II** – Companies must satisfy the requirements of all three pillars of Basel II, calculate and aggregate market, credit and operational risk measures, and minimize efforts and investments in risk management.

▫ **Business Process Outsourcing** - With proliferation of BPO, governments and businesses in Europe and the US are putting more pressure on India to strengthen measures to protect personal information.

▫ **California's SB 1386 Law** - Mandates that companies must report breaches of personal data to consumers. SB 1386 was the catalyst for data breach laws being considered in more than 30 states and enacted or are pending in Arkansas, Florida, Indiana, Montana, North Dakota and Wisconsin.

▫ **Consumer Privacy Protection Act of 2005 (H.R. 1263)** - Establishes certain rules on privacy notices to consumers, including privacy policy statements. Data custodians have certain statutory information security obligations.

▫ **COSO Risk Framework** - Originally formed in 1985 to sponsor the National Commission on Fraudulent Financial Reporting, which developed recommendations for public companies and their independent auditors, for the SEC and other regulators, and for educational institutions.

▫ **EU Data Protection Directive.** Establishes minimum data protection standards to be followed in the EU. Article 28 provides for the establishment of data protection supervisory authorities in each member state. Authorities have investigative and enforcement powers.

▫ **Freedom of Information Act 2000** - Gives people a general right of access to information held by or on behalf of public authorities, promoting a culture of openness and accountability across the public sector.

▫ **Health Insurance Portability and Accountability Act (HIPAA) 1996** - Calls for regulations promoting administrative simplification of healthcare transactions as well as regulations ensuring the privacy and security of patient information.

APPENDIX 2 CONTINUED: Regulatory Compliance Matrix

▣ **Information Protection and Security Act** - Directs the FTC to regulate information brokers. Data brokers are required to ensure data accuracy and confidentiality, authenticate and track users, detect and prevent unauthorized activity, and mitigate potential harm to individuals.

▣ **Information Protection and Security Act (S.500/H.R. 1080)** - Requires the Federal Trade Commission to regulate all "information brokers", which means virtually any business that maintains or processes personally identifiable data will be subject to the regulations.

▣ **Notification of Risk to Personal Data Act** - Requires people and federal agencies that own, license or collect personal information to notify individuals whose information was obtained by an unauthorized person.

▣ **Patriot Act 2001** - Anti-terrorism measure enabling the US Government to obtain access to a variety of information including personal data in the possession of businesses.

▣ **Personal Information Protection Law 2005** - Outlines a set of obligations that any company in Japan that holds personal data on 5,000 people or more must adhere to, including ensuring that personal data are kept secured and protected against unauthorized access and disclosure.

▣ **Sarbanes-Oxley** – Dictates that all companies must track, store, document and audit every financial process and control more effectively.

▣ **Social Security On-line Privacy Protection Act (H.R. 82)** - Establishes new FTC regulations for information brokers. Individuals have the right to obtain disclosure of all personally identifiable information pertaining to the individual held by an information broker, and to be informed of the identity of each entity that procured any personally identifiable information from the broker.

▣ **The Guidelines for Technical and Managerial Measures for the Protection of Personal Data** - These are binding guidelines for South Korean businesses to prevent data security breaches.

▣ **The Personal Information Protection and Electronic Documents Act 1998 (PIPEDA)** - Enacted to alleviate consumer concerns about privacy and to allow Canada's business community to compete in the global digital economy.

▣ **The Privacy Amendment (Private Sector) Bill 2000** – Drafted as an amendment to the Privacy Act 1988. Sets out how private sector organisations should collect, use and disclose, keep secure, and provide access to personal information.

FINANCIAL SERVICES

▣ **SEC 17a-3, 17a-4** – Amendments to SEC's broker-dealer books and records rules. The amendments clarify and expand recordkeeping requirements and expand the types of records that broker-dealers must maintain and require broker-dealers to maintain or promptly produce certain records at each office to which those records relate.

LEGAL ETHICS RULES

▣ **DR4-101** – US lawyers professional code of conduct that stipulates the preservation of confidence and secrets in electronic documentation.

APPENDIX 3: Types of Document Metadata and Their Associated Risks

Document metadata comes in many forms. Below is a list of the types of metadata found in Microsoft Office documents and the risks that each type of metadata poses to a corporation.

▫ **Document Properties** — Microsoft Word, Microsoft Excel, and Microsoft PowerPoint documents. Document properties are details about a file that help identify it that includes a descriptive title, subject, author, manager, company, category, keywords, comments, and hyperlink base. Document properties display information about a file to help organize the files so that they can be easily found at a later date.

   Risks — The names of authors and the name of the company can display sensitive information about a corporation. It is possible that if a document has been sent outside your own corporation, the author name and company name contained in the built-in properties could be a name other than your own. In addition, if documents are re-purposed or used as a template for a new document, information that is specific to a previous client such as pricing, terms, or the client's name can be stored as hidden information within the new document.

▫ **Document Statistics & File Dates** — Microsoft Word documents only. Document statistics include information on when the document was created, when it was modified, when it was accessed, and when it was printed. In addition, document statistics display the name of the person it was last saved by, the revision number, and the total editing time. Other statistics include number of pagers, paragraphs, lines, words, and characters.

   Risks — Document statistics can create embarrassing situations when the hours billed do not match the total editing time. In addition, the "last saved by" metadata shows the last person who edited the document. This can be risky if it is discovered that the person whose rate and time is billed out is different than the person who actually worked on the document.

▫ **Document Reviewers** — Microsoft Word documents only. Document reviewers consist of a list of users that have added or accepted any track changes. When the names of reviewers are removed, but not the Track Changes, the revisions remain with the document. However, the user name associated with each revision will be removed. It is recommended that the names of the document reviewers be removed when removing track changes.

   Risks — The risk from the Document Reviewers metadata is that it can expose who has suggested what changes.

▫ **Custom Properties** — Microsoft Word, Microsoft Excel, Microsoft PowerPoint documents. Custom Properties includes any property fields added manually to a document or by various programs to help manage and track files.

   Risks — Custom Properties are normally things specific to an organization. The potential risk arises because it is easy to see a history of this document.

▫ **Hidden Text** — Microsoft Word documents only. Hidden text are text blocks that have been formatted as hidden. Unless specifically selected to be viewed in Microsoft Word, hidden text is not displayed within the document.

   Risks — Hidden text can contain notes that are particular to a document. As hidden information that is not cleansed, the hidden text can potentially be viewed by unintentional parties.

▫ **Comments** — Microsoft Word, Microsoft Excel, Microsoft PowerPoint documents. Comments are notes and suggestions that are added to a document via the comment feature to help facilitate an online review.

   Risks — Comments, like hidden text, unless intentionally removed can display sensitive information to external parties because comment metadata travels with the document. Microsoft Excel and Microsoft PowerPoint documents are especially susceptible to this risk as there is no internal mechanism built into these applications to warn a user that comments are embedded. Gartner Group states that "while Microsoft is aware of potential problems [with comments], it does not have a comprehensive solution to solve this problem."

APPENDIX 3 CONTINUED: Types of Document Metadata and Their Associated Risks

▪ **Track Changes and Document Revisions** – Microsoft Word and Microsoft Excel documents. The Track Changes feature tracks changes (inserted, deleted, and moved text) made to a document during an online review. As changes are made to a document using Track Changes, a new revision of the document is kept by the application. This revision history exists, even after changes to the document have been accepted or rejected.

Risks – Track Changes shows the history of changes to the document. If Track Changes is left on, but the highlight on the screen is turned off, every change made to the document still remains. This is like recording every single keystroke made to the document that can be viewed by subsequent reviewers. Thus, even though the Track Changes are not visible, it still travels with the document and, in some circumstances, it can be sent to and seen by an unintentional party with potentially disastrous consequences.

▪ **Headers and Footers** – Microsoft Word, Microsoft Excel, Microsoft PowerPoint documents. Headers and footers are areas in the top and bottom margins of each page in a document. Text or graphics can be inserted in headers and footers—for example, page numbers, the date, a company logo, the document's title or file name, or the author's name—that are printed at the top or bottom of each page in a document.

Risks – Custom header and footers can contain descriptions such as filename, path, the date and time the document was modified, or other information that is deemed important to make it easy to retrieve and edit a file. Unfortunately, the information contained in footers and headers is often overlooked when the document is shared. Failure to remove this information can result in revealing confidential information.

▪ **Footnotes** – Microsoft Word documents only. Footnotes attributed to content are embedded as metadata into Microsoft Word documents.

Risks – Footnotes may expose private, internal directions about how the document is used in the organization.

▪ **White Text** – Microsoft Word documents only. White text is blocks of text that have been formatted with a font color of white on a background of white. The text appears invisible when viewed or printed and can be used to hide information in a document.

Risks – White text is commonly used when documents are posted to the Internet so that can be more readily found by search engines. However, white text can also be viewed by external users. Depending upon what was actually written as white text, the information can be very damaging. White text can also be used for particular field codes such as the "include text" field code, which can point to a file location. If this file location code is embedded in a document, users can unknowingly be updating the code and can potentially expose the document to a hacker.

▪ **Small Text** – Microsoft Word documents only. Any text block contained in a document that is less that five (5) points is considered small text. The text is so small that it will not be visible when viewed or printed and can be used to hide information in a document.

Risks – Like white text, small text is commonly used to put information in documents so they can be found by search engines. Small text can also include sensitive information that was not meant to be distributed externally.

▪ **Macros** - Microsoft Word documents only. If a task is repeated in Microsoft Word, it can be automated using a macro. A macro is a series of commands and instructions that are grouped together as a single command to accomplish a task automatically.

Risks - There are several reasons to strip out custom macros. For example, macros can be set for templates that may have some amount of pre-populated data. There may be a time when the information contained in these templates should not be seen by external audiences. Another example, macros can be linked to internal databases or intranets. Having access to the internal file naming structure is generally information that most corporations do not want outside their firewall. Lastly, macros are often quite complex and, if developed in-house, may represent the company's intellectual property. If macros are included in the document, the information is freely shared with any outside party.

APPENDIX 3 CONTINUED: Types of Document Metadata and Their Associated Risks

▪ **Previous Versions** – Microsoft Word documents only. Previous versions show the number of times that a document has been versioned over its lifetime. This function enables Microsoft Word to save prior versions of a document as a part of the electronic file.

Risks – The risk associated with previous versions is that a recipient can access any of the previous versions that have been saved. Therefore, the party reviewing the document can go back to any version and see what was changed in the document lifecycle. This metadata, while useful in some instances, can disclose sensitive information.

▪ **Routing Slips** - Microsoft Word and Microsoft Excel documents only. Routing slips are used to create a distribution list of reviewers in a particular order. Routing slips are manually created by adding in recipients' email addresses. When files are routed, it is sent as an attachment in an email message.

Risks - Routing slips reveal the names of the people that the document was sent to for review. This may be information that should stay confidential rather than distributed externally. An example of how this information can be used is when email addresses are put in the routing slips. If this document is then published to the Internet, the email address can be displayed for all to see.

▪ **Fast Saves** - *Microsoft Word documents only.* Fast saves is an option in Microsoft Word that saves just the changes that were made to a document, resulting in the history of the changes being saved with the document file. Turning fast saves off and saving the document will remove the changes and store only the final version of the document.

Risks - Like other metadata, changes saved during a fast save can expose sensitive information to external parties when viewed using a text or hex-editor. Deleted text can still exist in the electronic file. According to the Gartner Group's Research Note on Metadata in Office, "users can easily forget that metadata exists when they send the document to someone else. Some metadata is never visible, such as pieces deleted by users but not really deleted by Microsoft Office when operating with fast save turned on."

▪ **Hidden Slides** - *Microsoft PowerPoint documents only.* Hidden slides are slides that are hidden so that they are not shown during a slide show.

Risks - A master Microsoft PowerPoint slide deck may contain some slides that are uses as backup or that are for internal use only. To prevent accidental showing of these slides, it is best to strip out any hidden slides before sending the slide deck out externally.

▪ **Hyperlinks** - Microsoft Word and Microsoft Excel documents only. Documents can contain hyperlinks to other documents or Web pages and are displayed as blue underlined text. Hyperlinks in Microsoft Excel files can be seen in: a link to a cell in another Microsoft Excel document, a named link to a named reference in another Microsoft Excel document, a link to another document, an OLE link that inserts another document as an icon, and an OLE link that inserts another document as text.

Risks - Hyperlinks can maintain a link to a site that corporations may not wish to disseminate such as files that may exist on a computer's local file system, on a corporation's internal database, or on an intranet. Disclosing the file path, or the location of where the files are stored can invite potential hackers to gather sensitive corporate information.