Content and Metadata Leakage

BUSINESS VIEWPOINT

Securing Digital Content and It's Hidden Artifacts

Dan Keldsen
Senior Analyst, Consultant
dan.keldsen@delphigroup.com

November 2005

Traditional content creation and retrieval solutions want all of the latent metadata to remain within documents, as they are retrieval links for searching/browsing on content collections, but from a security perspective, these are mines hiding just below the surface of a seemingly innocuous landscape.

Due to the volume of content and speed of business, it is no longer reasonable to expect workers to be the watchful eye and enforcement point in cleansing this information.

DELPHI GROUP
A Perot Systems Company

111 Huntington Avenue, Suite 2750
Boston, MA 02199
(617) 247.1511

www.delphigroup.com

Over the last several decades, the business world has undergone nothing short of a revolution. Tools such as word processing, e-mail, IM and the Web have completely redefined the modern concept of communication. But, with all the benefits that have been afforded through these changes, there have also come new challenges with regards to security and compliance. The flood gates are cracking – something has to be done to align current communication approaches to risk management and corporate strategies. Virtually every modern organization must face the new challenges posed under the umbrella of content security.

For several years, Delphi has been advising it's consulting clients on the benefits (and dangers) of assuming that content is in any deep sense "secured" simply because an Enterprise Content Management (ECM) solution is in place (as one example), it has not been until the last 12-18 months that the constant headlines of not just identity theft (ID Theft) related content leakage, but also Intellectual Property (IP) leakage, or generically, "high-value" content have made these issues highly visible to the masses, and thus has brought this message beyond the early adopters/visionary organizations. Today's company's, whether through vision and strategy or in reaction to legal and government-based retribution are waking up to the fact that a proactive approach to securing content must be taken. At Delphi we call this approach Dynamic Information Access Control (DIAC).

While Delphi's work in the DIAC area for our consulting clients has been largely focused on the pro-active benefits of automating and understanding content-oriented systems (with of course an eye to compliance issues close at hand), respondents to our 2005 survey lean heavily towards more tactical than strategic issues. The richness and versatility of electronic content war-

rants such an approach. One must not only look at "protecting" the overt content, but the potential hidden content and metatags that can accompany an online file, as these can be as valuable or damaging as the content itself.

These digital artifacts which track multi-authorship of documents, prior versions of content, formatting, comments, and other aspects of "metadata" have no direct analog comparison from pre-computing eras. The full uses and abuses of this 'productivity enhancing' ability are only recently coming to being truly understood. This information, while generally of benefit to knowledge workers, librarians and record keepers, is also a potential danger if inadvertently exposed to the wrong set of eyes.

It is no longer reasonable to expect workers to be the watchful eye and enforcement point in cleansing this information. The gains of digital tools are rapidly being eroded due to the risks of these hidden artifacts, and to ignore this area of risk can be embarrassing at the least, to outright cause for corporate demise in the extreme. A proactive, policy-based omnipresent approach to protecting and securing content is mandatory for any prudent organization in the 21st century. That is the focus of DIAC.

Beyond (Traditional) Security - Focusing on Content

Organizations looking to secure their content must first carefully assess the needs of their high value and high risk content to determine what level of DIAC should be applied.

DIAC is security that specifically addresses – and in many cases is embedded into content throughout it's entire lifecycle and regardless of format or method of transmission. This policy-driven capability is distinct from the realm of Information Security (InfoSec), which is primarily focused on securing infrastructure such as networks, servers, desktops, and operating systems. DIAC includes not just entire documents of any file format, but can be extended down to "chunk" levels (such as chapters, paragraphs, indexes, etc.) as well as the metadata associated with this content, such as MS Office properties fields, tracked changes, forwarding history, etc.

Traditional content creation and retrieval solutions typically keep all of the latent metadata, as they are retrieval links for searching/browsing on content But from a security perspective, these are mines hiding just below the surface of a seemingly innocuous landscape. For example, final delivery of a response document to a Request for Proposals (RFP) document - it is unlikely that this hidden information would serve any other purpose other than to embarrass or damage the company sending their response. Separating the good from the potential harmful metadata is almost inevitably more trouble than it's worth, without adequate tools to assist.

DIAC ultimately allows content to be security aware **within context**

– such as who is accessing the content, where the user is currently located, whether the user is connected or off-line, how many times viewers are allowed to open the document, whether cut/copy/paste of content is allowable, whether rights to forward to another user are available, etc.. This particular set of capabilities is most fully embodied within Enterprise or Digital Rights Management (ERM/DRM) solutions. As with any security solution planning, it is not suitable to treat all content as equally valuable, so the temptation to jump straight to the most capable/complex implementation (such as ERM/DRM) wrapped around **all** content would be overkill on a massive scale. Organizations looking to secure their content must first carefully assess the needs of their high value and high risk content to determine what level of DIAC should be applied, to create a representative security scheme for content that neither creates undue burden (and expense) nor leaves content at risk inappropriately.

Why DIAC, Why Now?

From a technology view, DIAC capabilities have existed as point/niche solutions for several years. But like much technology-based functionality, sound business practices today mandate a more centralized, orchestrated, possibly single policy-based approach. This is the attitude that is driving the rise of interest in Information Architecture (IA) and Service-Oriented Architecture (SOA). Following this trend, DIAC solutions need to be addressed holistically through single integrated platforms, not silo-ed functions that potentially permit cracks in the security scheme.

We have found that organizations that address the larger issues of Intellectual Property Management (IPM) are capable of swiftly (and with minor modification to existing systems/policies), implementing regulation-specific policies.

From a business view, DIAC has risen in stature due to the increase in court decisions that turn on e-content resources and regulation and legislation of recent years, (e.g. Sarbanes-Oxley and HIPAA). It should be no surprise therefore, that Regulatory Compliance was identified as the number one business driver behind investments in DIAC, according to respondents to our 2005 DIAC survey.

However, while Regulatory Compliance tops the list, with a 28% response rate (respondents were limited to a single response on this question), we were greatly encouraged to see that Intellectual Property Management (IPM) ranked as the second highest response, at 18% - indicating the realization that Intellectual Property in and of itself should be managed with as much, and perhaps more scrutiny, than the more publicly noted concerns of customer or patient data/information.

In fact, we have found that organizations that address the larger issues of IPM are capable of swiftly (and with minor modification to existing sys-

In theory:

The recent visible rise in outsourcing is an area ripe for use of DIAC solution, employed to insure that Intellectual Property or sensitive data which is flowing through the outsourcing party remains within their systems, and does not find it's way, accidentally or purposefully, leaked to outsiders.

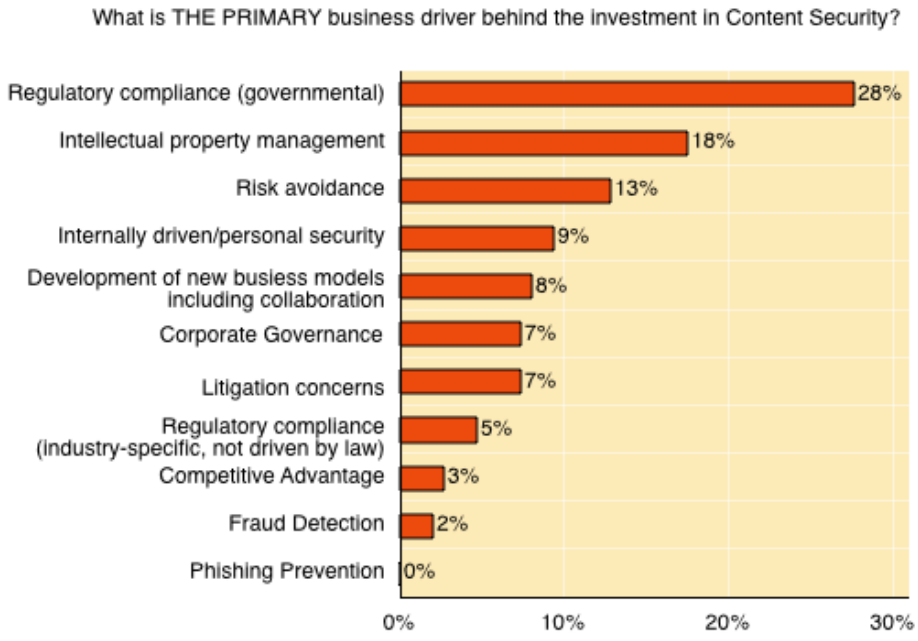
In reality:

Of buyers of such technology that we have interviewed, not a single one was actually using such a solution for this purpose. Until their competitors adopt such technology or customers demand it, outsourcers see no valid business reason to embrace this "solution."

tems/policies), implementing **regulation-specific** policies - and this in turn leads to even further strategic thinking around this set of capabilities.

Examples of more strategic capabilities than the typical entry point to DIAC include the ability to populate "virtual deal rooms" with context-aware security that enables companies to freely share content while a financing deal is on the table, but to instantly retract and destroy content when negotiations have stalled or the deal is no longer in negotiation.

Within the top five responses, just over 8% of respondents indicated that the "Development of New Business Models, Including Collaboration" was the **primary** business driver for investments in DIAC. On the surface, this level of response could easily be written off, however, this is quite a different mind-set to bring to investments in this area, as it focuses much more on beneficial aspects (revenue generation) rather than "yet another cost to minimize." This is clearly a minority opinion, but as the early adopters ramp up and innovate around this area, it is only a matter of time before this trend impacts the later adopters.



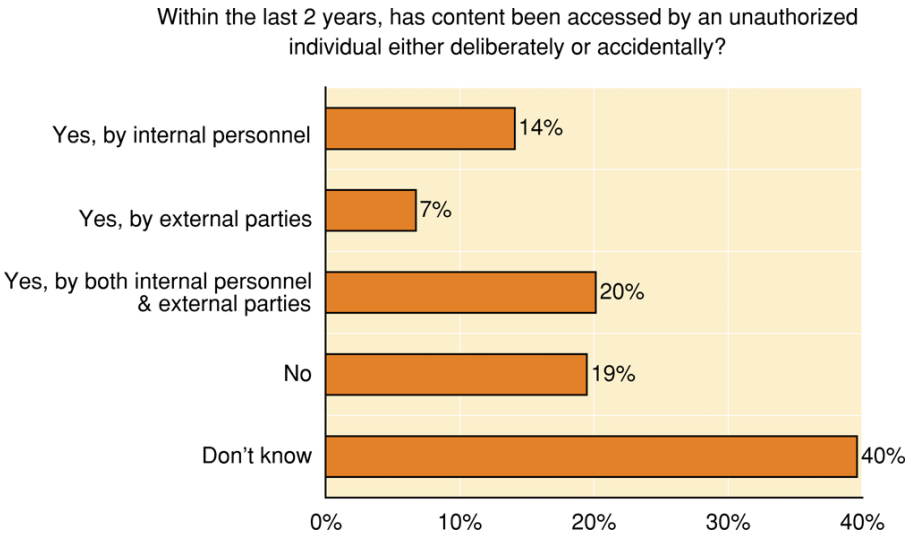
The temptation for many organizations to look at "complying with the letter of the law" will end up wasting both current and future money spent on redundant analysis and implementation above and beyond baseline compliance needs that are oriented towards more pro-active, or beneficial to the bottom- and top-line views.

Fully 40% of respondents companies' had no idea whether content had been accessed by an unauthorized person.

How Real are the Problems Around DIAC?

Since the "Dot Bomb" projects with no demonstrable business case(s) are very unlikely to be funded unless they are the pet project of an executive sponsor. DIAC is no different. In order to further discover what was driving interest in DIAC, one line of inquiries in our survey served to identify whether there was a discrepancy in how content was being secured against the potential threats, and whether existing policies and systems were serving organizations well in this regard, at the very least, in providing visibility into the depths of their information access violations.

Fully 40% of respondents companies' had no idea whether content had been accessed by an unauthorized person (or person abusing their existing access rights), which is a more honest assessment than most organizations would typically concede to, opting instead for "optimism" or outright denial. 41% of respondents said that internal, external or both internal and external parties had accessed content without authorization. Further, as many classic surveys (such as the oft-cited CSI/FBI surveys) have revealed over the years, the "insider threat" is again being cited twice as often (14%) as the external threat (7%) - such as hackers or crackers.



In total, 81% of respondents have had issues around content leakage/exposure, or have reason to be concerned, not the least of which, because they simply have no idea (via an audit trail or other notification method) what their level of exposure to this threat is, nor confidence in their ability to prevent, detect or react to these issues. A detail to keep in mind from this response is that respondents are speaking primarily of whole document or document collection leakage. The further subtlety of metadata leakage

In total, 81% of respondents have had issues around content leakage/ exposure... The further subtlety of metadata leakage would likely cause this response to move even closer to 90-95% of respondents experiencing content leakage from both directly visible content and hidden metadata.

would likely cause this response to move even closer to 90-95% of respondents experiencing content leakage from both directly visible content and hidden metadata.

Bottom Line

In today's litigious and regulated world, a world in which content is created and shared with potentially "anyone" anywhere in the world, the ease of pushing a few buttons, prudent organizations must ask themselves, what is our potential risk at not proactively and intelligent securing our content. Are we aware of what is posted in each file - both overtly and covertly in metadata and audit histories? Can we control how this content is shared throughout the entire lifecycle of the content. Can we administer sound and prudent security policy effectively, centrally and consistently? Unfortunately, the reality is today, most organizations are flying blind, completely unaware of the potentially damaging exposure that is already happening because of inadequate, antiquated approached to securing business content. These issues are lying in wait to be discovered by a competitor, regulator, disgruntled employee, hacker or newspaper reporters.

From an adoption standpoint, it is not realistically a question of whether an organization will begin to adopt DIAC capabilities, but **when, whether through strategy and vision - or reaction and recoiling**.

The clear path to the future deployment of DIAC solutions points to both more specific/targeted uses of these capabilities, and a push towards strategic vision from the executive offices, and execution from a technical management standpoint.

But... nirvana is not coming any time soon. There is certainly enough baseline capability available technically for DIAC-oriented solutions that the lack of "maturity" in this market should be taken with a grain of salt, and the serious business issues which are currently being ignored at least be raised as concerns to management. DIAC planning needs to be focused on the specific threats and opportunities that can be realized, and then addressing each one of these through application of available technology solutions, coupled with a sound business strategy and policy. The work can be daunting, but to ignore it could potentially lead to a far greater set of legal and financial issues.

Surveying the Landscape: Buyers and Suppliers of DIAC Capabilities

In 2005, Delphi Group ran a survey to assess perceptions, definitions, and current/expected deployments of DIAC solutions, the drivers for adoption (or lack thereof), exactly what content would be secured and from whom.

The resulting dataset of 458 responses came from a wide variety of industries, from Small-to-Medium Enterprises (SMEs) to Fortune 1000 organizations of sub $10M to >$25B Annual Revenue, predominantly headquartered within the United States.

Solution Examination - Workshare

As stated in the first half of this paper, the creation of DIAC platforms today requires the integration of different targeted technology solutions. The second half of this paper focuses on one such technology from Workshare - a technology component that directly addresses the management and security issues related to the "hidden" risks in content as well as the monitoring of content as it is being shared/sent, in order to "catch" potentially confidential/inappropriate content from being shared.

Workshare's roots are in solutions created to facilitate efficient & accurate document review through the document lifecycle oriented primarily around MS Office document formats. As anyone who has distributed copies of Office files to a number of colleagues, and then needed to reconcile all suggested changes, even using track changes, knows that this is a complex, and frequently manual task. As a result of the functionality Workshare provides in this area, they became intimate with the inner workings of both the visible and invisible (metadata) content within Office documents. This knowledge led to the realization that there are other business issues associated with online file creation, beyond the authoring process itself.

This approach proactively removes the burden from the individual to "do the right thing" in sanitizing documents appropriately, and simply make it a matter of policy, - a centrally enforced policy.

That is why Workshare expanded beyond their original functionality, to include capabilities to manage hidden metadata created as a result of single or multi-person authorship, in a policy-driven, manner. This approach proactively removes the burden from the individual to "do the right thing" in sanitizing documents appropriately, and simply make it a matter of policy, - a centrally enforced policy. The product was further enhanced to provide content-based and policy-based monitoring of e-mail and attachments to determine if the content being sent (potentially) violates security, compliance or corporate policy rules.

Workshare provides this functionality via a suite of software offerings, These security-oriented capabilities are such functions as: Manage Document Rights, Security Policy Management, PDF Security Options, Security Notifications, Document Risk Report, and Assembly of a Secure Master file from disparate separate documents, authored/reviewed in a collaborative fashion. The product offerings include:

Workshare PROFESSIONAL - the full suite of capabilities within one product, from collaborative authoring/reviewing facilitation to automatic security and distribution of finalized documents.

Workshare DELTAVIEW and Workshare DELTAVIEW PE - reconciliation of collaborative authoring/reviewing, still actively maintained and revised along with entire suite.

First Glimpse of Workshare Hygiene:

As mentioned in the overview of Workshare's offerings, they have been focused largely on securing/managing hidden data (metadata) historically.

As work on this whitepaper was wrapping up, Workshare announced at the DEMOfall 2005 event their "Document Hygiene Technology." With this, Workshare is moving squarely even deeper into the DIAC world as we had described in the beginning of this whitepaper.

With their "Hygiene" technology, they are now offering policy-driven capabilities surrounding the document lifecycle across **both** hidden (metadata) and visible content to manage, audit, block outright or "cure" (remove the "disease" causing a policy violation, via Workshare's MS Office integration capabilities) content within the enterprise, via centrally created, yet de-centralized enforcement points.

Initially, this "Hygiene" technology is provided within Workshare Professional 4.5 (as of October 2005), and is slated to be expanded to provide network-side capabilities in Q1 2006.

Workshare PROTECT - which provides only the security and audit capabilities found in Workshare PROFESSIONAL. For those employees who have no need for the collaboration aspect, this offering still allows for automatic policy enforcement, assisting the employee in automatically producing documents "to code" - with metadata stripped, automatically rendered as PDF and secured to only allow on-screen viewing, with a specific password, for example.

TRACE! - a freeware offering, provides a taste of the awareness capabilities of Workshare PROFESSIONAL or PROTECT. TRACE! reports on what metadata is buried within documents, keyword matching against common concerns such as profanity, potential inclusion of Social Security Numbers in documents, and so on. This tool allows an individual to gaze into the current risk levels within existing document collections on a local machine, web server, and within stored e-mail.

As of the wrap-up of this paper, Workshare had announced an extension of security & compliance capabilities via their "Document Hygiene Technology" (see sidebar for an early glimpse).

When deployed in a deliberate manner as part of a DIAC strategy, these tools can make a positive impact on the ability of an organization to minimize the risk their content represents, while simultaneously increasing the value derived from this content through collaboration.

To illustrate how this can be accomplished, Delphi examined two current users of Workshare products, to gain not just a theoretical basis for the role of Workshare in DIAC, but real world appreciation and application.

Case Study: LeBoeuf, Lamb, Greene, & MacRae

LeBoeuf, Lamb, Greene, & MacRae (www.llgm.com) is a law firm founded in 1929, of approximately 1400 employees, with 12 Domestic (US) offices and 5 major International offices. Their path through the concerns of metadata and inappropriate information leakage echos that of Workshare's product expansion from document comparison ("red-lining") to metadata management & cleansing and securing of documents.

LLGM came to be aware of Workshare having needed a replacement for Lexis-Nexis' CompareRite product - an offering scheduled for end of life by Lexis-Nexis in 2002. As they adopted the Workshare DeltaView solution, their relationship with Workshare and Workshare's offerings began to further blossom. LLGM has since adopted Workshare Protect as well (early 2004), for a more complete solution. Their standard desktop (hard drive) image integrates includes the Windows OS, full Office Suite, includ-

ing MS Exchange as well as Workshare DeltaView and Protect on all 1400 desktops.

As a Law Firm, there is really no choice but to have a dedicated red-lining solution outside of MS Office's "Track Changes" capabilities, as that is not a terribly robust internal capability. The further step of removing metadata traces of all of the reconciliation activity and other inadvertent content that may of concern from a privacy and corporate compliance perspective, is most definitely not addressed "out of the box" by the MS Office suite. In LLGM's search for a metadata/document compliance solution, they evaluated Workshare Protect against other offerings, and found that Workshare Protect was by far the easiest to install, manage/support, train and use than any other solutions they had examined.

Organizationally there was some initial concern about the imposition of policies from a central location (via Workshare Protect) on the various partner's practice areas, but they have been flexible in their deployment of policies/capabilities. Their policies provide assisted document cleansing/protection, with ultimate responsibility for appropriate communications being handled at the desktop by the individual. With this solution in place, they are looking to revisit this to look at more completely automating security policy imposition on final documents and ultimate distribution method to their clients.

Case Study: Buzzacott

Buzzacott is a mid-tier accounting firm of 150 employees and 20 partners based in London. They provide general professional accounting services, as well as specific services for charities, and expatriate taxation advice. Their business is based on high-quality, high-trust relationships with their clients, including not only expertise in their advice and work done on behalf of their clients, but in ensuring that confidential information remains confidential with close attention to detail in who receives final documents and the state of contents therein.

Buzzacott's implementation is based on Workshare Protect, and is used to ensure that their internal quality and compliance standards are met without undue manual oversight being required. Prior to using Protect to facilitate the workflow process of document assembly, metadata stripping, PDF conversion and securing according to policy, they found that manual processing was causing not only significant delays in scouring documents from top to bottom, but inevitable human error in not completely following company policy to the letter.

DELPHI GROUP
A Perot Systems Company
111 Huntington Avenue, Suite 2750
Boston, MA 02199
(617) 247.1511
www.delphigroup.com