ITM Whitepaper

ILLINOIS INSTITUTE OF TECHNOLOGY School of Applied Technology

...because knowledge is power.

Controlling the Emerging Data Dilemma:

Building Policy for Unstructured Data Access

Anne Shultz
Edited by Ray Trygstad

12/23/09

www.itm.iit.edu

An Information Technology and Management Whitepaper from
Illinois Institute of Technology's School of Applied Technology

ILLINOIS INSTITUTE OF TECHNOLOGY *School of Applied Technology*

ITM Whitepaper: *Building Policy for Unstructured Data Access*

It's everywhere. It's saved blatantly on the desktop of a coworker's unattended computer, just waiting to hop onto the next flash drive and head out of the company. It lingers just a click away, ready to be uploaded and emailed to a competing company. It lies nakedly on a manager's desk, eager to be picked up by criminal hands. It lurks in an unsecured network drive, hoping to be discovered by someone with malicious intentions. It's unstructured data and it's demanding attention.

What is Unstructured Data?

In general, unstructured data can be defined as any electronic information without a specific structure. Depending on the context, this definition can indicate data which is stored outside of a database as well as documents where the contents can take any shape, much like the text in a Word document. This includes documents, blueprints, presentations, image files, video files, and so on. However, it is important to remember that whether or not the data is considered structured depends on the context. For example, although spreadsheet data can be structured in cells and arranged in rows and columns, like those created with Excel, this is not controlled by the application[1]. For this reason, spreadsheets should be considered unstructured data.

Merrill Lynch estimates that unstructured data makes up over *85 percent* of all business information[2]. To make matters worse, the amount of unstructured data within companies is still growing. With email and file services being the biggest contributors, more and more information is becoming available electronically and easy to share[3]. According to a study by the Aberdeen Group, a yearly increase in the amount of unstructured data generated throughout the organization was reported by 86% respondents[4]. As it comprises such a large percentage of business information, one would assume that management of unstructured data and unstructured data access would be a priority for most organizations, but a survey developed by the Ponemon Institute and Varonis System Inc. indicates differently. According to this study, which surveyed 870 IT operations professionals, 91% of organizations do not have a process for establishing ownership of unstructured data[5]. Further, 76% of respondents were not able to determine who can access unstructured data, while almost 70% felt that employees in their organization had unnecessary access to unstructured data[6]. Lastly, 89% of respondents to this survey acknowledged that controlling access to unstructured data is very difficult for their organization[7].

Why is Unstructured Data Access a Problem?

The looming beast of unstructured data is a serious issue for companies from a legal standpoint. Businesses lacking control over unstructured data access may be ill prepared when it comes to legal discovery. In the event of a lawsuit, all related documents must be held as potential evidence. If there is no control over unstructured data in general, required documents may be difficult to find in the time allotted by the court[8]. Searching for documents may be challenging if it has not been determined who is responsible for the information. Further, "chain-of-custody" must be verified for any documents held during the litigation process, and to verify chain-of-custody, a company must prove that the documents are authentic and are what they claim to be[9]. This means there must be documented proof of when the documents were created, who they were created by, what was done with the documents, and who accessed or viewed the documents. Verifying chain-of-custody may prove to be nearly impossible with no control over unstructured data access. In addition to main-

[1] Dorian, P. (2007, March). *FAQs: unstructured data FAQ.* Retrieved from http://searchstorage.techtarget.com/guide/faq/category/0,,sid5_tax306615_idx0_off10,00.html

[2] Atre, S., & Blumberg, R. (2003, February). "The problem with unstructured data." *Information Management Magazine*, February 1, 2003. Retrieved from http://www.information-management.com/issues/20030201/6287-1.html

[3] *Ibid.* 1.

[4] Aberdeen Group (2009, July). *Securing unstructured data: How best-in-class companies manage to serve and protect.* Retrieved from http://www.nymity.com/Free_Privacy_Resources/Previews/ ReferencePreview.aspx?guid=d7c2 b604-3f7e-491a-90f4-c2db075a5613

[5] StorageNewsLetter.com (Ed.). (2008, July 1). *Organizations lack control of their unstructured data assets* [Press release]. Retrieved from http://www.storagenewsletter.com/news/miscellaneous/varonis-ponemon-institute-unstructured-data

[6] *Ibid.*

[7] *Ibid.*

[8] *Ibid.* 1.

[9] Murchison, R. S. (2009). *Retention management for consistency & compliance* [PowerPoint slides]. Available from http://www.matchps.com/training.html

taining chain-of-custody, any retention policy mandated by the company will be difficult to enforce if it has not been determined who is accountable for maintaining the data. If the retention policy is not applied evenly, documents may be deleted prematurely or kept longer than the retention policy requires. Either of these situations will point to an inconsistent retention policy and could cause serious trouble for a company faced with providing documents in a court of law.

Lack of control over unstructured data is also a problem for businesses when it comes to compliance. In light of today's corporate compliance requirements, such as Sarbanes-Oxley, PCI, and HIPAA, many businesses must tighten controls on their processes and systems. This also involves tightening controls for systems which handle unstructured data. For example, the Sarbanes-Oxley act requires strong access controls to ensure that financial information is not corrupted[10]. This includes strong access controls for financial systems, as well as for unstructured financial data. The Payment Card Industry (PCI) Data Security Standards also require strong access controls in order to ensure sufficient protection for customer credit card information. PCI requirement 7.2 maintains that access to cardholder information must be denied for all employees unless access is absolutely needed for their job[11]. Like Sarbanes-Oxley, this rule applies not only to credit card systems, but to unstructured credit card data as well. Yet another act which requires tighter access controls around unstructured data is the Health Insurance Portability and Accountability Act (HIPAA). HIPAA sets security standards in order to maintain "confidentiality and integrity of individual health information"[12]. These security standards require strong access control over any information systems which handle individual health information, including those which handle unstructured information, such as file systems[13]. In general, without the ability to control unstructured data in general, a business will find any of these compliance controls difficult to meet

In addition to any legal and compliance implications, a lack of control over unstructured data access is also a problem from a general security and productivity standpoint. As the Ponemon Institute and Varonis survey demonstrated, 76% of respondents were not able to determine who can access unstructured data and nearly 70% of respondents felt that employees in their organization had unnecessary access to unstructured data[14]. With no control over the access to unstructured data, highly confidential or sensitive information could easily fall into the wrong hands and possibly leak to the public. Depending what type of information is leaked, this could impact the company's ability to be competitive in its dealings or even damage the company's ability to do business.

Whether the lack of control over unstructured data access is a problem for legal and compliance reasons or simply general security reasons, it is obviously something that needs to be done. The good news is that more and more solutions are surfacing in the area of unstructured data. The bad news is that none of these "solutions" seem to have completely solved the problem. Throughout this paper, we will review available methods for controlling unstructured data access and propose a strategy for developing a foundation for Unstructured Data Access Policy.

Available Unstructured Data Access Solutions

When considering any problem related to information, it is typical for businesses to look first to technology solutions. This idea seems to have held true for unstructured data access as well. As concern about this problem has gained momentum, more and more technology solutions have surfaced with the promise to improve organization and productivity. These technologies have many different names but can generally be referred to as content management systems or document and record management systems. For the sake of simplicity, throughout this paper they will be referred to as content management systems or CMS. Content management systems can be used as unstructured data repositories which allow the information to be organized and controlled. Basic components of

[10] Lambert, L. K. (2009, February 4). *Access management and SOX compliance*. Retrieved from http://www.securityinfowatch.com/root+level/1296049

[11] Burton Jr., J. D., Chuvakin, A., Elberg, A., Freedman, B., King, D., Paladino, S., & Shcooping, P. (2007). *PCI compliance: Implementing effective PCI data security standards*. Burlington, MA: Syngress Publishing.

[12] Infotechadvisor. (n.d.). *HIPAA: Comprehensive guide*. Retrieved from http://trygstad.rice.iit.edu:8000/HIPAA/HIPAA%20Guide%20Part%20I%20-%20infotechadvisor.mht

[13] *Ibid*.

[14] *Ibid*. 5.

content management systems include document repository, integration with desktop applications, check-in and check-out, versioning, auditing, classification and indexing, and search and retrieval, and security[15].

The most relevant component of content management systems in the context of unstructured data access is security. Azad Adam explains that "Security should be tightly integrated with the system, allowing for security access permissions to be applied at different levels within the system"[16]. An adequate content management system may allow security to be assigned to groups or individuals as well as to groups of documents or individual documents. For example, an administrator should have the ability to assign one group of users the ability to read and edit a specific document while assigning another group of users the ability to read the document only. Still another group of users may not have access to see that specific document at all. As another example, an administrator should be able to assign access to a specific folder such that all users have the ability to read documents stored in the folder while only one user has the ability to edit them.

Content management systems handle access differently with unique options for securing data at multiple levels of granularity. An example of a content management system with growing popularity is Microsoft Office SharePoint. SharePoint users can be granted access in two ways. First, as with most content management systems, access permissions can be granted to a user or group of users[17]. Second, SharePoint makes use of collaboration sites which are essentially websites used to organize display groups of documents[18]. By way of inherited permissions, collaboration sites can be used to allow for more creation with less access management overhead. If a group of users have read-only access to a collaboration site and the site is configured to inherit permissions, that same group of users will have read-only access to all subsequent sites as well[19]. Another example of a content management system with unique security capabilities is Laserfiche. In addition to offering security permissions at a group or individual user level, Laserfiche also allows users to control access to specific documents through the use of security tags[20]. For example, if a user is assigned to a security tag titled "Confidential," that user will have access to see documents that have the "Confidential" tag applied to them. Further, if that user is creating or saving a document in Laserfiche, they will have the ability to apply the "Confidential" tag to their own documents. These are just two examples of the many diverse content management systems available. Regardless of the specific functionalities offered by the software, any content management system will no doubt propose a unique solution to the problem of unstructured data access.

At first take, it seems that content management systems should be the perfect solution to the problem of unstructured data access. However, contrary to the claims of CMS vendors, this is not likely to be the case. The fundamental issue with content management systems lies in the establishment of policy. In other words, these content management systems cannot be used effectively if it is not first established how access should be configured. The authors of *Integrative Document & Content Management* explain that development, communication, and acceptance of a policy framework should be completed before even beginning requirement specifications for a content management system[21]. To further emphasize this, the authors state, "the development of a policy framework is not dependent on an investment in [content management systems]. The policy framework can be developed to apply improved practices for managing documents using existing tools"[22]. This is an extremely important point for any technology solution. A policy must be established first to support

[15] Adam, A. (2008). *Implementing electronic document and record management systems.* Boca Raton, FL: Auerbach Publications.

[16] *Ibid.*

[17] Curry, B., English, B. (2008). *Microsoft Office SharePoint Server 2007 best practices.* Redmond, WA: Microsoft Press.

[18] Microsoft (2007). Microsoft Office SharePoint Server (Version 2007) [Software]. Available from Microsoft: http://sharepoint.microsoft.com/how-to-buy/Pages/default.aspx

[19] *Ibid.* 17.

[20] Laserfiche (2008). Laserfiche 8 (Version 8.0) [Software]. Available from Datanet Solutions: http://www.datanet-solutions.com/content/enterprise-content-management.html

[21] Asprey, L., & Middleton, M. (2003). *Integrative document & content management: Strategies for exploiting enterprise knowledge.* Hershey, PA: IGI Global.

[22] *Ibid.*

the needs and fundamental requirements of the organization. Only after the policy has been developed and accepted should an organization turn to a technology solution for the possible automation of controls required by the policy.

Further, since the policy serves as a cornerstone for any technology configurations, it is imperative to realize that the effectiveness of the content management system depends on the effectiveness of the policy. As an example, to ensure that the importance of policy is taken into account for SharePoint deployment, the following is stated in *Microsoft Office SharePoint Server 2007 Best Practices*: "SharePoint Server 2007 provides a multitude of security features which, when implemented in concert with *well-understood information security policies*, provide significant protection of confidential information"[23] (emphasis added). This statement clearly emphasizes that security functions offered in the software are only effective when implemented with comprehensive policy. This is an important basis of information for any content management system deployment which brings about another dilemma; what makes an effective policy for unstructured data access?

Guidance for building Unstructured Data Access Policy for use with content management systems is still lacking. Furthermore, until it is established, content management systems may never be used effectively for unstructured data access. However, it is helpful to remember that the wheel does not need to be completely re-invented in this situation. Standard access structuring methods have already been developed and may prove useful if applied to unstructured data. Examples of these methods include discretionary access control, mandatory access control, and role-based access control, and attribute-based access control[24].

First, discretionary access control (DAC) is based on "ownership of information," and "delegation of rights". In a DAC model the creator of the information is also considered the owner and administrator of the information. This means that the creator is responsible for granting or revoking access to their information. It also means that the creator has the ability to grant administrative access to other users so that they may grant or revoke access to the information as well. Compared to other access control model categories, DAC is considered to be the most simple. Because of its simplicity, there are many security conditions are not taken into account wen using DAC. For example, since users are responsible for granting and revoking access, any security requirements are also the responsibility of the users and cannot be easily managed by organizational authorities. Another condition unaccounted for is the possibility of cascading revocation chains where one user removes access from someone immediately after the access has been granted by a different user. Most fundamental, however, is the lack of control over the flow of information with DAC. Access can be granted whenever and to whomever at the discretion of the user.[25]

Next, mandatory access control (MAC) is centered around the idea that access is set up based on predefined, mandatory rules. There is no notion of ownership involved with MAC. Instead, in order to access information protected by MAC, the user must possess the appropriate security clearance required for accessing the information[26]. MAC strategies are considered to be "lattice-based access-control systems"[27]. The flow of information in a lattice based access-control system is predetermined by the mathematical structure of that specific model. Many MAC models have been developed including need-to-know, Bell-LaPadula, and Biba[28,29]. Each MAC model uses a specific mathematical formula to govern how access will be structured. The development of the MAC concept was driven by policies created for military environments. While these static control methods work well in hierarchical military context, they are typically too rigid for use in enterprise organizations. This is due to the fact that MAC controls cannot be changed unless amended by an administrative authority and thus do not permit sharing of information across the organization.[30]

[23] *Ibid*. 17.

[24] Lopez, J., Furnell, S. M., Katsikas, S., & Patel, A. (2008). *Securing information and communications systems: Principles, technologies, and applications.* Norwood, MA: Artech House.

[25] *Ibid.*

[26] Benantar, M. (2006). *Access control systems: Security, identity management and trust models.* New York, NY: Springer.

[27] *Ibid.*

[28] *Ibid*. 24.

[29] *Ibid*. 26.

[30] *Ibid*. 26.

Role-based access control (RBAC) is used most widely in large enterprise organizations and originates from the concept of grouping users by job function[31]. The idea is that users who share the same job functions will require similar access rights[32]. For example, a role would be created with the access permissions required by a particular job function and that role would be granted to all users performing that job. Two categories of RBAC include Hierarchical RBAC and Constraint RBAC. Hierarchical RBAC is the idea of roles having an order based on access levels. With hierarchical RBAC, roles may be inherited to acquire the access permissions of lesser or greater roles. Constraint RBAC is an RBAC concept used to accommodate segregation of duty constraints required and can be accomplished through static or dynamic separation of duties. First, static separation of duty is the method of ensuring segregation of duties by using a separate role for each job function. Dynamic separation of duty is the method in which roles are only activated if the situation permits. For any role-base access control method, any users joining the organization, changing job functions, or leaving the organization must be accounted for and have their access updated accordingly. In addition, roles must be reviewed and updated on a regular basis to catch any access rights which may need to be removed or added to any particular role.[33,34]

Finally, attribute-based access control (ABAC) was developed to accommodate security requirements of larger, dispersed systems. In this method, access is determined by user attributes and is granted in a role style, similar to RBAC. Examples of attributes could be position, department, age, location, etc. Depending on an individual user's specific attributes, they will be granted specific predetermined access permissions. The idea is that a user's access permissions will be changed as their attributes change. For example, if a user changes departments, they will be granted access specific to their new department while access specific to their old department will be removed. It is important to note that although ABAC is considerably more flexible, this method does imply greater complexity in the creation and maintenance of policy.[35]

The fact that access structuring methods have already been established is comforting in light of the unstructured data access dilemma. However, the question now remains; Why haven't these access structuring methods been applied successfully to enterprise unstructured data? Although these methods exist, there is no guidance available to help organizations decide which method will work best. In the context of unstructured data, how is a business to decide which structure should be used or how their information should be organized within that structure? Although access structuring methods have been established, understanding on how to effectively incorporate these methods into an Unstructured Data Access Policy is still lacking.

Fortunately, there is one emerging concept which appears to be filling the position as the next fundamental puzzle piece in the development of Unstructured Data Access Policy. This is the concept of concept of *Data Governance*. With Data Governance, organizations are learning to step back and develop data access strategies from an enterprise point of view. Gwen Thomas, of The Data Governance Institute, describes data governance as, "a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods"[36]. More simply, data governance can be considered as the concept of making decisions about what should be done with information. Most importantly, Data Governance promotes the idea that security of information is no longer the sole responsibility of the Information Technology department, but that it should involve the enterprise as a whole. The organization of information security must be conceived with the entire organization in mind so that rules and access philosophies are applied consistently throughout. To explain this fundamental Data Governance concept, Thomas uses the analogy that Information Technology is like a plumbing system with pipes, pumps, and

[31] *Ibid.* 24.
[32] *Ibid.* 26.
[33] *Ibid.* 24.
[34] *Ibid.* 26.
[35] *Ibid.* 24.
[36] Thomas, G. (n.d.). *The DGI data governance framework.* P. 3 Retrieved from http://datagovernance.com/dgi_framework.pdf

storage tanks. Thomas explains further that, "data is like the water flowing through those pipes". Using this analogy, the goal of Data Governance is to addresses issues specific to what is "flowing through the pipes". In order to address these issues, input is required from management and subject matter experts who understand the data—those who control the spigots—and not from the "plumbers" of the system.[37]

Depending on the goals of the company, Data Governance projects may focus on different areas, or may even focus on 2 or 3 areas at once. These different categories of Data Governance focuses include policy/standards/strategy, data quality, privacy/compliance/security, architecture/integration, data warehouses/business intelligence, and management support. The Data Governance focus which sets the stage most effectively for the creation of unstructured data policy is that which focuses on privacy, compliance, and security. This type of Data Governance program usually originates from data privacy or access management concerns. New regulatory compliance, contractual, or internal requirements may also play a role in inspiring a program of this type. Thomas explains that a Data Governance program with this emphasis is likely to include initiatives for a number of tasks focused on securing the information. One of these tasks would be to support the use of access management and security requirements to safeguard sensitive data. Assisting the risk assessment and development of risk management controls is one more task in a Data Governance program of this nature. Another task involved with securing information is to ensure the enforcement of compliance requirements. Aligning initiatives and frameworks will also be done as part of such a program, along with identifying stakeholders, determining decision privileges, and clarifying responsibilities.[38]

It is important to realize that although initiatives for a Data Governance program may be similar, the details of the program will be specific to the organization. The purpose of Data Governance is to understand the information in the context of the organization and develop a method for governing the information based on the specific needs of that enterprise. Due to the subjective nature of the project, outcomes and deliverables of a Data Governance project may differ from organization to organization.

Proposed Solution for Unstructured Data Access Policy

At this point, we would like to introduce a strategy for using Data Governance in combination with access structuring methods to develop a foundation for Unstructured Data Access Policy. In this proposed method, Data Governance can be utilized as an essential prerequisite for the development of Unstructured Data Access Policy. While Data Governance deliverables will vary depending on the goals of the organization, we believe that some deliverables are crucial if the organization has plans to establish an adequate policy for unstructured data access. The first step in this proposed strategy is to ensure that essential Data Governance deliverables have been completed sufficiently. Deliverables should include an established governing body for information security related matters, a document retention schedule, clear establishment of information owned by each department, as well as clear sensitivity handling levels and procedures.

To explain the essential deliverables further, the governing body, which should be established as part of the Data Governance program, must include at least one knowledgeable representative from each department. These representatives will be needed to lead the development of unstructured data access procedures within their own departments. An Unstructured Data Access Policy lead must also be established to be the overall organizer of the policy. This individual should be well versed in the Data Governance project as a whole, and should also be capable of leading and organizing subject matter experts from each department in order to facilitate the development of unstructured data access structures and practices for the entire enterprise. It is important to remember that the role of the Unstructured Data Access Policy lead as well as the roles of the department subject matter experts are ongoing. These positions will be necessary not only for the creation of the policy but for the continuous maintenance of the policy as well.

[37] *Ibid.* Pp. 3-5.

[38] *Ibid.* Pp. 7-9.

Next, a document retention schedule must be established as part of the Data Governance program. While the actual retention periods specified on the schedule will not be used directly for the development of an Unstructured Data Access Policy, the categories of data content listed on the schedule are extremely important. Since the effectiveness of the Unstructured Data Access Policy depends on the accuracy of the content categories, it is of utmost importance that these categories be considered thoroughly. Content categories should accurately reflect all types of information content handled at the organization[39]. However, it is also important that content categories be broad enough that they will not need to be constantly modified; in general, categories should represent types of data content found in the organization and should also overlap to some degree with main business processes found in the organization[40]. The document retention schedule on the whole should be well-understood throughout the company with content categories easily identified by all employees.

Clear establishment of information owned and used by each department is another important deliverable which should be used in the development of unstructured data access. This may be added as an addendum to the document retention schedule, and should essentially list 3 groups of content categories for each department. These groups should include content categories owned by that specific department, owned by all departments, and used by that specific department. A department which owns a content category should also be the department responsible for retaining or destroying documents within that content category according to the document retention schedule. Content categories owned and used by a department can be distinguished by considering "whose lap the document falls into"[41]. For example, suppose a Human Resources department creates a business case showing estimated costs for a specific project. Once the business case is approved, perhaps it is standard practice within the company for the Purchasing department to keep the final, signed copy for budgeting purposes. In this case, the business case "falls into the lap" of the Purchasing department. In this situation, Purchasing would be the owners of the business case content category while the Human Resources department would simply use the category. Still, other documents seem to fall into every department's lap and should be considered as being owned by all departments. These include documents such as polices, forms, as well as other non-department specific types. An example of content category groupings for an Information Technology (IT) department within an organization may resemble that listed in figure 1.

Figure 1. Example of content category groupings for an IT department.

Owned – Department Specific:

- System Development Documents
- System Maintenance Documents

Owned – Department Non-Specific:

- Form Masters, Templates
- Policies, Procedures, Manuals
- Research, Reference Materials
- Projects, Subject Matter Working Files
- Calendars, Appointment Books
- Training Class Educational Materials, Handouts

Used – Owned by Another Department:

- Organizational Charts, Employee Lists (owned by Human Resources)
- Personnel Files (owned by Human Resources)
- Budgets & Forecasts (owned by Accounting)
- Business Cases, Vendor Bids, Proposals, Quotations (owned by Purchasing)
- Audit Final Reports, Collateral Workpapers (owned by Internal Controls)
- System Monitoring, Access, Audit Trails (owned by Internal Controls)

Finally, in order to develop a comprehensive Unstructured Data Access Policy, clear sensitivity handling levels and procedures should be established as part of the Data Governance program. These are levels of classification used to define the sensitivity of documents as well as how such documents should be handled as part of a data classification model[42]. Well understood sensitivity hand-

[39] Murchison, R. S. (January 2009). Personal communication.

[40] *Ibid.*

[41] *Ibid.*

[42] Whitman, M. E., & Mattord, H. J. (2008). *Management of information security, Second Edition.* p. 270. Boston, MA: Thomson Course Technology.

ling levels are necessary to guide employees on the acceptable use of confidential information within the organization, facilitating the proper use of an unstructured data access structure. One simple example of a data classification model could include levels such as Public, For official use only, Sensitive, and Classified[43]. In this type of model, specification for "Public" documents would be any document acceptable for release to the public, such as a press release. "For official use only" may indicate documents that are not especially sensitive but that should be kept within the organization, such as internal communications. "Sensitive" documents may signify documents which are considered to hold important information, potentially embarrassing the company or damaging market share if leaked to the public. Lastly, "Classified" information may indicate extremely confidential information which could significantly harm the interests of the company[44].

Once the Data Governance program has been launched and all deliverables essential for development of Unstructured Data Access Policy have been completed, the second step of policy development can begin. The second step is to determine all access situations that must be accounted for. Each department should complete this task individually but by using the same methods. Rather than beginning with known access structuring methods such as DAC, MAC, RBAC, or ABAC, we suggest looking at the actual information and understanding what type of security each type of document actually requires. Determining all access requirements can be accomplished by asking the following questions: What data needs to be accessed by who? For how long? A matrix may be helpful in determining access situations. A matrix example that we would like to suggest for determining what data needs to be accessed by who is one which examines the information in from both vertical and horizontal access planes within the organization. In this type of matrix, management levels of the organization can be listed vertically while departmental groupings can be listed horizontally as in figure 2. Throughout this paper, this matrix will be referred to as an Access Requirement Matrix.

Figure 2. Access Requirement Matrix demonstrating vertical and horizontal access levels.

Horizontal Access Levels

Vertical Access Levels

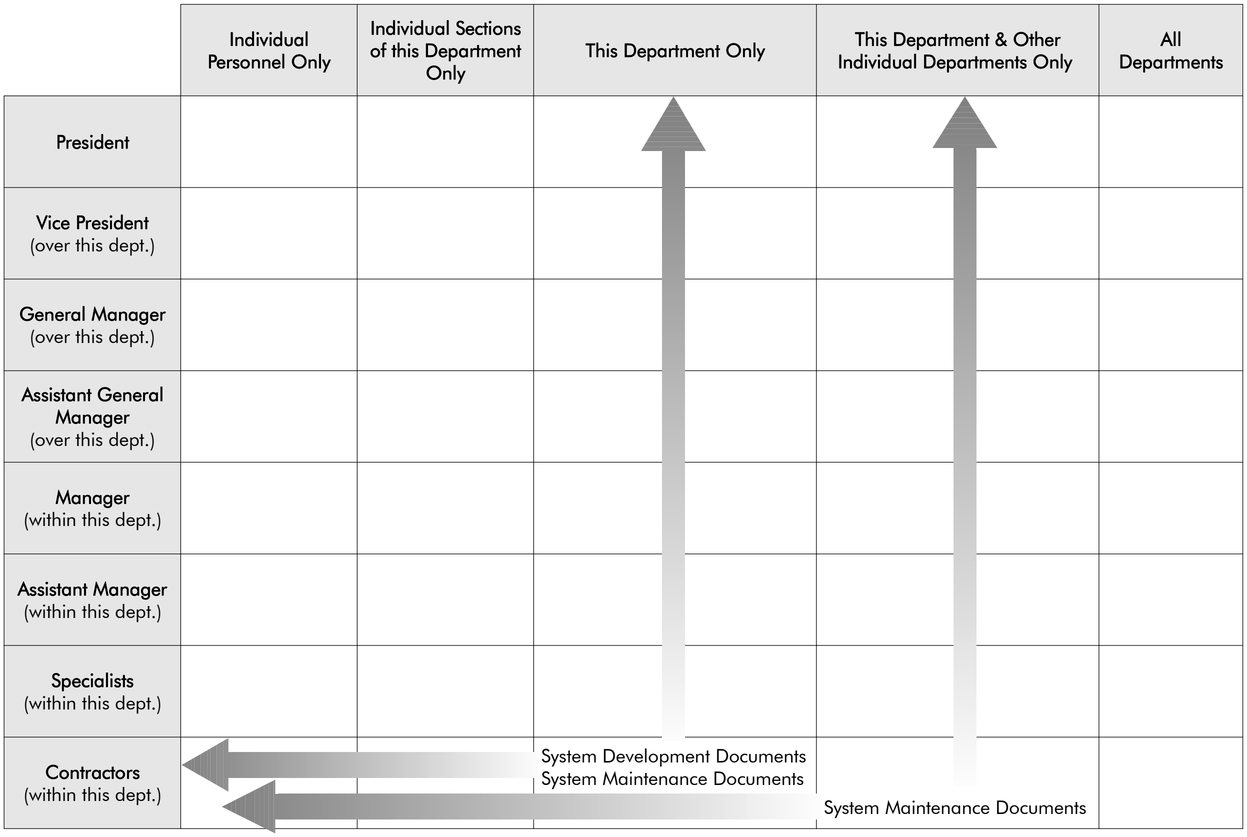
	Individual Personnel Only	Individual Sections of this Department Only	This Department Only	This Department & Other Individual Departments Only	All Departments
President					
Vice President (over this dept.)					
General Manager (over this dept.)					
Assistant General Manager (over this dept.)					
Manager (within this dept.)					
Assistant Manager (within this dept.)					
Specialists (within this dept.)					
Contractors (within this dept.)					

[43] *Ibid.* P. 271.

[44] *Ibid.* P. 271.

Using an Access Requirement Matrix, content categories can be mapped according to the access levels they require, similar to an ABAC method in which access is determined by attributes. Of course, the vertical and horizontal access levels listed in the matrix will vary, depending on the organization structure of the company. To limit complexity, a separate Access Requirement Matrix should be used for each department. To demonstrate this concept, we will use the IT department model from Figure 1. Remember that the content categories owned by the example IT department included System Development Documents and System Maintenance Documents. Suppose that both of these content categories are shared between all sections and all individuals within the IT department. Further, suppose it is acceptable for all levels within and above the department to access these categories. Suppose that in addition to System Development and System Maintenance Documents that are shared within the department, System Maintenance Documents also exist which must be accessed by employees from other departments. For example, documents used for maintaining financial applications may need to be accessed by employees from the organization's Finance department. To satisfy both of these access requirements, System Development and System Maintenance documents could be placed in the matrix as illustrated in figure 3. The gray arrows in figure 2 have been added to demonstrate the vertical and horizontal levels which should have access to these content categories, according to the placement on the matrix.

Figure 3. Access Requirement Matrix demonstrating content category access for System Development and System Maintenance documents.



	Individual Personnel Only	Individual Sections of this Department Only	This Department Only	This Department & Other Individual Departments Only	All Departments
President					
Vice President (over this dept.)					
General Manager (over this dept.)					
Assistant General Manager (over this dept.)					
Manager (within this dept.)					
Assistant Manager (within this dept.)					
Specialists (within this dept.)					
Contractors (within this dept.)			System Development Documents System Maintenance Documents	System Maintenance Documents	

In order to complete an Access Requirement Matrix for a department, each of the content categories should be entered into the appropriate cell according to the access requirements for documents within the category. A completed Access Requirement Matrix for the exemplified IT department may look similar to that illustrated in figure 4 on the next page.

Figure 4. Access Requirement Matrix using IT department example.

	Individual Personnel Only	Individual Sections of this Department Only	This Department Only	This Department & Other Individual Departments Only	All Departments
President					
Vice President (over this dept.)					
General Manager (over this dept.)					
Assistant General Manager (over this dept.)					
Manager (within this dept.)				Personnel Files Budgets & Forecasts	
Assistant Manager (within this dept.)					
Specialists (within this dept.)				Audit Final Reports, Collateral Workpapers System Monitoring, Access Audit Trails	
Contractors (within this dept.)			System Development Documents System Maintenance Documents Form Masters, Templates Policies, Procedures, Manuals Research, Reference Materials Projects, Subject Matter Working Files Calendars, Appointment Books Training Class Educational Materials, Handouts	System Maintenance Documents Organizational Charts, Employee Lists Business Cases, Vendor Bids, Proposals, Quotes	Form Masters, Templates Policies, Procedures Manuals Training Class Educational Materials, Handouts

Once an Access Requirement Matrix has been completed, each cell of the matrix indicates a different unstructured data access configuration which must be accounted for in the Unstructured Data Access Policy. However, it is important to understand that this matrix is developed primarily for determining access requirements which do not change often, similar to an MAC method. Since access requirements established through use of an Access Requirement Matrix are considered primarily static, additional processes may need to be considered to account for possible temporary situations. One common example of a temporary access situation would be that of a cross-departmental project. In this situation, non-standard groups of individuals may require access to particular documents during the life of the project. A suggested method of handling this type of situation is to create a process in which the project manager is responsible for determining the access. In this case, the project manager would determine which individuals should have access to the project file as well as when the access should expire. This type of process may resemble a DAC method in which a user is responsible for granting and removing access. However, for temporary situations such as projects, it is important that the start and end dates be respected. Once a project comes to a definite end, all documents to be retained according to the retention schedule should be moved from the temporary project file to appropriate locations based on content category.

Once unstructured data access requirements have been developed, procedures must be developed for the granting, revoking, and changing of access. Whether it is possible for these processes to be automated or whether they must be monitored and completed manually, it is vital that these procedures be developed and documented. Without establishment of these procedures, unstructured data access will be unenforceable and the policy will quickly become ineffective. In addition, it is critical that each department review their unstructured data access requirements on a regular basis to ensure that needs are still being met. If any changes have occurred to the content categories or the organizational structure, Access Requirement Matrices must be updated accordingly. Any file access configurations developed from the Access Requirement Matrices must also be updated accordingly.

Finally, after all unstructured data access requirements, processes, and procedures have been developed, the writing of the policy may begin. Many effective policy writing guides exist and can be applied to the development of an unstructured data policy. Whitman and Mattord offer a framework for an issue-specific policy type which could be used for the development of an Unstructured Data Access Policy; use of an issue-specific framework should allow for the Unstructured Data Access Policy to roll up to a general Information Security Policy for the enterprise[45]. This framework includes Statement of Purpose, Authorized Uses, Prohibited Uses, Systems Management, Violations of Policy, Policy Review and Modification, and Limitations of Liability. In order to create an Unstructured Data Access Policy, the content of these sections should be based on the processes and requirements established specifically for the control over unstructured data access. This will include requirements specified through the Access Requirement Matrices or any other processes developed to account for temporary access situations. Processes and procedures developed for the granting, revoking, and changing of access should also be included as a foundation for an Unstructured Data Access Policy.

The strategy for Unstructured Data Access Policy foundation provided in this paper has been developed using research in combination with experiences gained through implementation of a Data Governance program at New United Motors Manufacturing Inc. (NUMMI). NUMMI is a privately owned auto manufacturing plant located in Fremont, California. It is important to note that the strategy proposed has not been formally tested or evaluated. Future areas of research could incorporate assessment of this strategy along with other Unstructured Data Access Policy solutions. Methods for effectively managing unstructured data access after it has been established could also be included in future research. In general, unstructured data access control methods must be developed further if organizations hope to harness the full potential of unstructured data. Unstructured data has become extremely easy to share and controls must be established to ensure that sharing is done appropriately within an enterprise setting. Further, with growing legal, compliance, and security issues, unstructured data will only become a greater issue if not addressed. With solid development of unstructured data access management and policy, this information may finally receive the attention it deserves.

ABOUT THE AUTHORS

Ann Shultz is a graduate student in the Information Technology and Management degree program at Illinois Institute of Technology. She was previously employed at New United Motors Manufacturing Inc. in Fremont, California where she held a position in Information Security compliance and was deeply involved in the design and implementation of their Data Governance program.

Ray Trygstad is the Director of Information Technology for Illinois Institute of Technology's School of Applied Technology and the Associate Director of IIT's Degree Programs in Information Technology and Management. He teaches courses in open-source operating systems; operating system virtualization; multimedia; information systems security management; and incident response, disaster recover and business continuity. He also teaches in the Master of Public Administration program at IIT's Stuart School of Business.

[45] *Ibid.* P. 119.

References:

Aberdeen Group (2009, July). *Securing unstructured data: How best-in-class companies manage to serve and protect.* Retrieved from http://www.nymity.com/Free_Privacy_Resources/Previews/ReferencePreview.aspx?guid=d7c2 b604-3f7e-491a-90f4-c2db075a5613

Adam, A. (2008). *Implementing electronic document and record management systems.* Boca Raton, FL: Auerbach Publications.

Asprey, L., & Middleton, M. (2003). *Integrative document & content management: Strategies for exploiting enterprise knowledge.* Hershey, PA: IGI Global.

Atre, S., & Blumberg, R. (2003, February). "The problem with unstructured data." *Information Management Magazine*, February 1, 2003. Retrieved from http://www.information-management.com/issues/20030201/6287-1.html

Benantar, M. (2006). *Access control systems: Security, identity management and trust models.* New York, NY: Springer.

Burton Jr., J. D., Chuvakin, A., Elberg, A., Freedman, B., King, D., Paladino, S., & Shcooping, P. (2007). *PCI compliance: Implementing effective PCI data security standards.* Burlington, MA: Syngress Publishing.

Curry, B., English, B. (2008). *Microsoft Office SharePoint Server 2007 best practices.* Redmond, WA: Microsoft Press.

Dorian, P. (2007, March). *FAQs: unstructured data FAQ.* Retrieved from http://searchstorage.techtarget.com/guide/faq/category/0,,sid5_tax306615_idx0_off10,00.html

Infotechadvisor. (n.d.). *HIPAA: Comprehensive guide.* Retrieved from http://trygstad.rice.iit.edu:8000/HIPAA/HIPAA%20Guide%20Part%20I%20-%20infotechadvisor.mht

Lambert, L. K. (2009, February 4). *Access management and SOX compliance.* Retrieved from http://www.securityinfowatch.com/root+level/1296049

Laserfiche (2008). Laserfiche 8 (Version 8.0) [Software]. Available from Datanet Solutions: http://www.datanet-solutions.com/content/enterprise-content-management.html

Lopez, J., Furnell, S. M., Katsikas, S., & Patel, A. (2008). *Securing information and communications systems: Principles, technologies, and applications.* Norwood, MA: Artech House.

Microsoft (2007). Microsoft Office SharePoint Server (Version 2007) [Software]. Available from Microsoft: http://sharepoint.microsoft.com/how-to-buy/Pages/default.aspx

Murchison, R. S. (2009). "Retention management for consistency & compliance" [PowerPoint slides]. Available from http://www.matchps.com/training.html

StorageNewsLetter.com (Ed.). (2008, July 1). "Organizations lack control of their unstructured data assets" [Press release]. Retrieved from http://www.storagenewsletter.com/news/miscellaneous/varonis-ponemon-institute-unstructured-data

Thomas, G. (n.d.). *The DGI data governance framework.* Retrieved from http://datagovernance.com/dgi_framework.pdf

Whitman, M. E., & Mattord, H. J. (2008). *Management of information security, Second Edition.* Boston, MA: Thomson Course Technology.

ABOUT ILLINOIS INSTITUTE OF TECHNOLOGY'S SCHOOL OF APPLIED TECHNOLOGY

Illinois Institute of Technology's School of Applied Technology offers hands-on, project-based technology-oriented education and training for both full-time students and working professionals. Courses are taught by IIT professors and industry professionals with significant working, teaching and research experience in their fields. The School of Applied Technology offers degree, non-degree, certificate, credit, non-credit programs, corporate training, short courses and seminars ranging from a few hours to several days in length. Both Bachelors and Masters Degrees are offered in Information Technology & Management and Industrial Technology & Management, as well as Undergraduate Certificates in Industrial Technology & Management, Graduate Certificates in Information Technology & Management topics and adult education/CEU courses in all fields. Our Information Technology & Management curriculum is supported by extensive dedicated laboratory facilities. We offer an comprehensive range of courses in information security and business continuity.

For more information on our education and training programs in information technology, please see http://www.iit.edu/cpd/

Illinois Institute of Technology (IIT) is a private, Ph.D. granting university founded in Chicago in 1890, offering programs in engineering, science, technology, architecture, design, psychology, public administration, technical communication, business and law.