boldonjames
A QinetiQ Company

30 YEARS 1985 - 2015

WHITE PAPER

DATA CLASSIFICATION: DELIVERING
SECURITY AND BUSINESS VALUE

INCLUDING

- THE INFORMATION MANAGEMENT CHALLENGE
- INFORMATION SECURITY DRIVEN BY BUSINESS VALUE
- REGULATORY COMPLIANCE
- IMPLEMENTING A CLASSIFICATION SOLUTION

INTRODUCTION

Information is critical to the business of most organisations. From design details for manufacturing to witness statements for criminal investigations, the right information allows an organisation to be successful. In the current business environment, much of that critical information is stored in digital form on computer systems.

In some cases this information is stored in databases. Databases, by their very nature, are highly organised storage systems that allow for easy management. Finding one piece of data in a structured database is easily accomplished with software tools that can automatically index, sort and search.

Other forms of information storage, namely documents and email messages, do not have the inherent structure of rows and columns that database tables have. It is harder for software to arrange and locate the bits of information required. In turn, it is harder for users to carry out the tasks that depend on finding and using information.

One method of reducing the difficulty of this task is by classifying information into categories. Categories can range from as simple as the department responsible for the information to as complex as national security classifications. Identifying the type of information allows both users and IT departments to properly manage an organisation's information. For users, information classification provides guidance on how they should handle information. IT departments can use the same information classification to tailor IT infrastructure to effectively manage that information.

DATA CLASSIFICATION
DELIVERING VALUE

WHITE PAPER

THE INFORMATION MANAGEMENT CHALLENGE

Anyone who has worked in a modern business environment knows the challenges of trying to find the information they need in the sea of documents and emails that every organisation generates. From desktop folders and personal mailbox stores to shared network drives and document management solutions, an organisation's information is spread across a wide range of platforms and services.

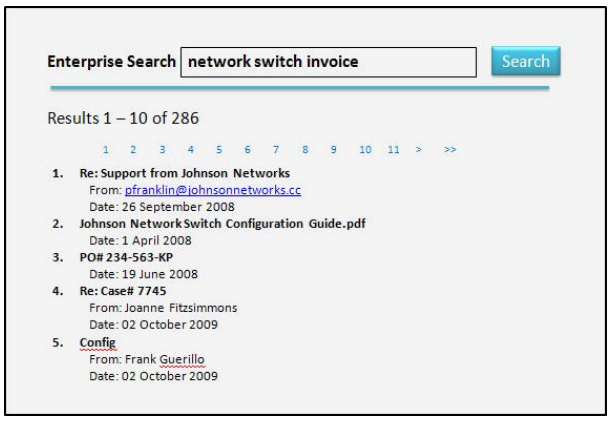
Of course, IT departments recognise the failures in some of these methods and implement services and policies to help increase the effectiveness for users and reduce costs. For example the use of workstation-based personal folders is discouraged so that backup and archive procedures can effectively protect an organisation's data. In general the move towards more centralised storage and management of data has meant better control over information.

But not all problems are solved by collecting the mass of business information inside the data centre. Users are now faced with large volumes of documents to search. The very nature of centralised information stores means that people are presented with more documents that are less relevant to the task at hand. For example, a payments administrator looking for invoice details is uninterested in the IT network configuration diagrams. To help reduce the number of files users have to search through to find their information, IT departments usually implement a folder structure that compartmentalises documents into business areas. A typical network file share might have an Accounting folder, a Human Resources folder and a folder for each employee – each with its own set of sub-folders. This filing system works. But only if everyone follows the rules and no one makes mistakes.

Mistakes and misuse are common and generally not unexpected. IT departments are asked to help business units clean up their files and folders on a regular basis as information becomes disorganised. There are IT solutions that can be implemented to help users track down lost documents. One common approach is to use document management software and a search engine. Document management solutions, by themselves, are no better at managing information than a network file share system. Documents will still get misfiled, effectively lost to the organisation. And for some users, hiding documents is standard practice. To help corral information, many organisations implement a search tool.

There are enterprise search tools that can peer into desktop files, network file shares, document management repositories, email message archives and backup tapes. The indexes produced by search tools can be extensive enough to cover all the digital information an organisation has ever produced. However, a search engine's results are hit and miss if users don't have the exact keywords that are in the documents they are looking for. Adding too many search words leaves the user with no results, adding too few returns too many.

The problem with search engines is that they lack a context for the information that is indexed. An email message about a payment invoice is weighed the same as a message containing a joke about invoices. For the most part, documents and email messages don't contain this context in a way that search engines can effectively make sense of information. There are some search engines that allow a user to specify keywords that appear "near" each other, or return results that exclude keywords. But the search engine cannot know what the information is about when reading the body of an email message or the contents of a document.



Information classification is one method of helping users and IT services to understand the context for a piece of information. For example, knowing that a particular email message contains a joke about an invoice means a user can move on in the search for a specific email message about an invoice. Often called categorisation or cataloguing, information classification adds to the content of digital data. This extra information could be the name of the business unit responsible for the document, or the retention period required by policy for this type of email message. In the IT world, information classification data added to documents and email messages is called metadata ("data about data" or "information about information").

DATA CLASSIFICATION
DELIVERING VALUE

WHITE PAPER

Metadata is used in many diverse situations, from the memo line on cheques to ISBN numbers for books. In everyday conversations, words like “project plan” and “financial data” are used to describe the information that people need to do the business of an organisation. Although there are some terms that are common to many businesses, each organisation has its own way of describing the information that it depends on. In some larger organisations there are even terms that are appropriate for only part of the business. These words and phrases can be thought of as the taxonomy for the information classification.

EXAMPLE OF A SIMPLE TAXONOMY

DEPARTMENT
Accounting
HR
Logistics
Service Programmes
Senior Management

Returning to the invoice search example above, if the user was able to narrow the search to only those email messages that were related to the accounting department, the chance of success would increase and the time required to conclude the search would decrease.

In some organisations, the complexity of the business can warrant extending the taxonomy that describes information. Additional fields or types of categories can be used to provide more details about documents. For example, one category might be the department of responsibility for information, and a second category could be the associated fiscal year. In government departments, information is often required to include a security classification and description.

Information classification allows users and IT services to add ‘context’ to the raw information that is stored in documents and email messages. The context for the data provides the opportunity for efficiencies in business processes and information management. Not only can a user more easily find information needed to perform tasks, but the IT infrastructure can also take advantage of metadata to reduce costs.

Most IT departments run archiving and backup solutions to protect the information that an organisations relies on. However, as mentioned earlier, not all information stored by an organisation is of equal value. There are types of documents and email messages that are essential to the normal functioning of a business. And there are documents that are of no real value to that same business. And there are some documents whose value lies somewhere in between the two extremes. Without an effective method of differentiating between these types of documents an IT department must treat all content with the same level of care. This approach ends up wasting resources by spending too much time and money on useless data and not enough on critical data.

One advantage to using an information classification strategy for an organisation’s information assets that might not be as obvious is reduced cost during e-discovery procedures. As long as an organisation can credibly state that information has been categorised, the magnitude of an e-discovery process can be limited to the information that is relevant to the issues at hand. Reducing the number of documents and email messages involved directly reduces the cost of an e-discovery exercise.

INFORMATION SECURITY DRIVEN BY BUSINESS VALUE

One area of an IT department that can benefit substantially from information classification is security. The addition of metadata describing the type of document or email message can inform IT security services about what policy is appropriate for that piece of data.

From a security perspective, there are numerous types of data an organisation needs to manage (for example, financial details, trade secrets and customer data). Each kind of data will have a specific set of policies that govern processing, storage and disclosure. For example, most jurisdictions have laws and regulations covering the disclosure of personal identification details. The people who are allowed to see and process these types of data are usually well defined and well known within an organisation. What isn’t well known is the set of documents and email messages that are subject to the organisation’s various security policies.

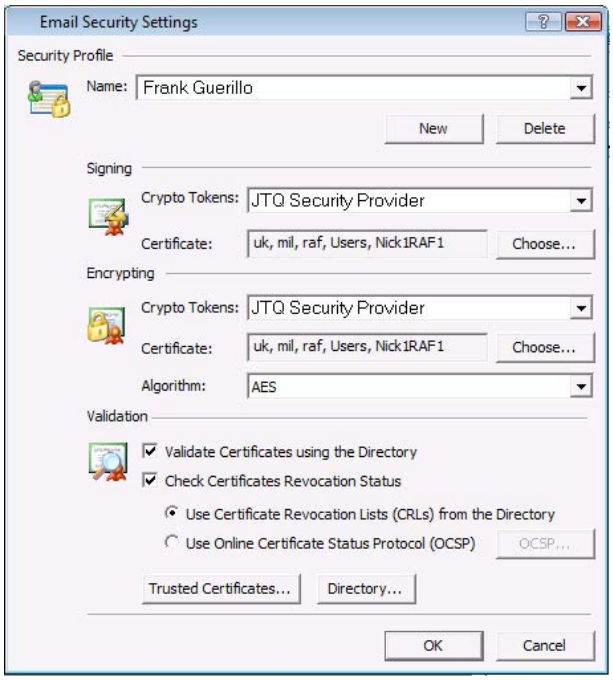
IT solutions, such as encryption and access control measures, are very good at maintaining the security of data once it has been protected. But knowing which of the many documents and email messages should be protected is a problem and protecting all the content at the highest level is impractical for most organisations. Add to this the difficulty users have in understanding which policies to implement, on what content, with which security tool and inevitably data is left unprotected.

Metadata can assist IT security within an organisation by addressing both the ‘ease of use’ and the ‘applicability’ problems. It is difficult for most users to understand the implications of choosing one security option over another when handling documents and email messages. What users do understand is the value that their actions and the information they are handling have to the organisation’s business. By using information classification that links the business value of documents and email messages to IT security policies, a user can be responsible for the security of information without having to understand the detail of the IT security software.

DATA CLASSIFICATION
DELIVERING VALUE

WHITE PAPER

For example, a common security measure used with email messages is encryption of the content for the intended recipients. The problem of knowing when an email message must be encrypted is usually defined in a policy document that a user might read once a year. The problems of invoking encryption range from remembering to click the toolbar button, to searching directories for certificates. The challenges for an average user can be daunting.



By adding additional information to an email message based on the value the content has to the business, the IT security policies for that specific type of information can be applied. For example, an email message regarding customer account details can be automatically protected by most email encryption tools. In this way, business terminology can drive the application of security solutions.

METADATA DRIVING YOUR SECURITY SOLUTIONS

There are a number of common IT services that can take advantage of information classification metadata added to documents and email messages. In addition to document management (and by extension, records management), enterprise search, archive and email encryption, metadata can be used to enhance the effectiveness of existing IT infrastructure including Data Loss Protection (DLP) solutions, Electronic Document and Records Management Systems (EDRMS), and email gateways.

DATA LEAK PREVENTION (DLP)

DLP can provide an effective layer of security to an organisation's information. DLP engines work, for the most part, the same way that search engines do. By matching terms and phrases found in documents and email messages, DLP engines assign a score to content. This score is used to evaluate the content as users and systems copy and move documents and email messages in and out of an IT network (some DLP vendors include a categorisation element that assigns metadata to the content). When an email message is sent, or a document is moved or copied, the DLP solution evaluates the DLP score against defined policies and then reacts (forbidding the action, warning the user or logging the activity).

However, most DLP implementations suffer from two serious problems. First, tuning the DLP engines to the profile of actual business content can be a very long and expensive process. Most DLP vendors recommend scanning all content and then analysing the results to find patterns that reflect the way the organisation does business. This first step in deploying DLP is usually measured in months. And as the organisation grows and changes, these DLP configuration steps need to be updated to match the changing business.

The second problem that DLP solutions have is inherent in the architecture of the software. Most DLP architectures rely on software that lives outside the content creation process, residing as software in the data centre, or at best as a service running on desktops. The creation and management of information and the evaluation of policy are disconnected. This results in two flaws. First, when DLP software makes a mistake there is no person accountable for the mistake. A malicious user cannot be held accountable for the errors in configuration of a DLP system. Second, the intention of a user cannot be respected. At no point does a user get a chance to apply a DLP policy to the content that might not match the DLP analysis of that content. For example, if an organisation's top financial manager wants to force an email message to be treated as business confidential she has no way to ensure that the DLP solution treats it as such.

By adding information classification metadata to documents and email messages, and configuring DLP policy to understand the value of this metadata, an IT organisation can significantly increase the effectiveness of a DLP implementation. Metadata contained in an email message indicating the sensitivity of the content can be recognised by a DLP solution and the correct policy applied based on the user's intention without guesses based on DLP software algorithms.

ELECTRONIC DOCUMENT AND RECORDS MANAGEMENT SYSTEM

Records management is a serious challenge for many government departments and regulated industries. Information management regulation must be adhered to in all cases, or individuals and organisations risk significant penalties. Even for general business, EDRMS can be a useful tool to help manage an organisation's ever growing quantity of digital information.

DATA CLASSIFICATION
DELIVERING VALUE

WHITE PAPER

Information classification can be used to help users manage documents and email messages that need to be entered into an EDRMS. The addition of a metadata flag that indicates whether a piece of content is a record helps both users and IT services identify a file that should be entered into the EDRMS. Depending on an organisation's workflow processes, an automated system could locate documents and email messages from central storage locations and upload the results to the records store. This method of flagging content is especially useful when email messages are created.

EMAIL GATEWAY

Email is the number one business communication tool in most organisations. We all rely on being able to transmit messages to co-workers, partners and customers. Protecting the email service is an important job for IT departments. Installing an email content scanning gateway allows organisations to check email messages for malicious code, spam, over-sized messages and inappropriate content, providing a cleaner and more robust tool for users.

Email gateways can be configured to inspect not only the content of a message, but also the metadata information. By building policy rules an IT department can provide added security. For example, an email gateway might be configured to block any messages destined for outside the organisation that does not have a metadata field indicating it is approved for release.

REGULATORY COMPLIANCE

Some organisations have a strong regulatory requirement for adding information classification to their business processes. Government departments in most jurisdictions (for example UK, USA, Australia, European Union) are required to add visual markings to documents and emails that clearly display the classification of the contained information. Users have a very strong incentive to add these visual marks to every document and email message as a matter of national security.

The security classification of documents is generally related to the level of impact that disclosure of the information would have to the security of individuals, commerce and the nation. Most governments also relate the security of information to the policies that govern information handling and IT systems. For example, a government may require that RESTRICTED information not be transmitted across public networks. There are many ways to implement IT security policies, but most methods require some knowledge about the content that is being processed in order to be effective.

IMPLEMENTING A CLASSIFICATION SOLUTION

There are a number of benefits that information classification can bring to an organisation, but there is an important prerequisite to make it work effectively. Consistent application of information classification is as vital for users as it is for the IT systems that depend on the metadata. The two aspects of this consistency are the standardization of the taxonomy and the application to all information. Adding classifications to some documents and not to others leaves gaps in the application of IT services to the content.

If only some of an organisation's documents are categorised, users may become confused as to when information classification is required and IT services can only implement policies on content that has the extra metadata. Having uncontrolled values for the classification of information lessens the effectiveness of those IT services. If the set of categories for information classification is left open-ended, there can be no consistent application of IT policy on content.

Having a comprehensive set of policies regarding user behaviour in the area of information classification is insufficient to garner its benefits. Users may have the best of intentions, but inevitably the use of information classification will be regarded as a hindrance to accomplishing the work of an organisation. Users need stronger incentives to encourage the adoption of new policies and procedures.

Using a software solution to handle information classification is a step in the right direction for a successful implementation. There are some characteristics of classification software that can help to ensure that an implementation is effective:

- **User Awareness:** User interaction in the process of applying information classification helps in a number of ways. First, the user becomes aware of the process. An aware user is more likely to make the right decisions when handling sensitive information than a user who does not know that information classification is used. Second, a user involved in the creation of the content of a document or email message is in a better position to understand the nature of the information and hence the applicable category for the information classification. Users are much better at understanding context than computer software.
- **Centralised Policy Management:** A second characteristic of information classification software is that it be centrally managed by the IT department. If left to the whim of the user, any software that could be reconfigured will be reconfigured. An information classification solution must be strictly controlled by the organisation that owns the information that is categorised. Organisational changes are often reflected in the information classification policies of a business. These changes need to be thoroughly planned, tested and rolled out across an organisation backed by policy.
- **User Experience:** The software that is chosen should be as seamless and familiar to the user as possible. The process of choosing a category for a document or email message should seem to the user to be an integral part of their normal suite of desktop applications. A user should need little or no training on how to use the information classification software.
- **Document Format Support:** An organisation that implements an information classification software solution should also consider the breadth of content that needs to be covered. If documents are stored in multiple formats, the software should be able to add information classification metadata to as many of the formats as possible. Additionally,

if different methods are used to generate content (for example, a desktop email client and a web-based email client) the software solution should be integrated with each of these tools.

- **Standardised Metadata:** The method of storing the information classification metadata in the content files should be done in a standard method. The reason for this is so that third-party software and IT services (such as security software or document management software) can read the categorisation information and trigger appropriate policies for managing the documents and email messages.
- **Policy Enforcement:** If an information classification solution is to be used for security purposes, the software should include a policy engine that can restrict or warn users when their actions violate an organisation's rules. For example, an organisation may not allow a user to downgrade the security classification of a document or email message. The software should be able to detect such an attempt and based on policy, warn the user or prevent the action but always log the event. The software should also provide a centralised tool for creating and distributing the policy to all users.
- **Audit Log:** The information classification software should create an audit log that contains the details of any interactions a user has with content. For example, the user's name, the name of the document or subject of the email message, the categorisation information, the attempted action and the result. This audit log can then be used when tracing a security incident and to provide a measure of compliance to the organisation's policies.

BOLDON JAMES DATA CLASSIFIER

Building on thirty years of experience in highly secure environments for the military and intelligence communities, Boldon James has created a suite of information classification plug-ins for the Microsoft Office platform. Boldon James Email Classifier and Office Classifier are developed to integrate seamlessly into the Office user interface for 2003, 2007, 2010 and 2013 editions and provide users with a simple method of adding visual and metadata information classification. With a centralised administration model that leverages Active Directory, an organisation can easily deploy software and policy across the user community.

USER INTERFACE

Boldon James Email Classifier allows an organisation to add information classification details to each email message. Users are provided with a toolbar or Ribbon Group for selecting labels from up to three categories. For example, one organisation may use one category for the sensitivity of the information (e.g. PUBLIC, INTERNAL, BOARD MEMBERS), whereas another organisation may require all three categories (e.g. Classification, Department, and Project). The administration tool, allows the centralised policy to define the rules that automatically apply visual tags to email messages. The labels chosen by the user can be added to the subject field of the email message and to the body of the message.

Boldon James Office Classifier is a labelling interface for Word, PowerPoint and Excel. Users are presented with an interface that is consistent with Boldon James Email Classifier, reducing the requirements for additional training. Marking policy can define visual marks in a document's header, footer and/or watermark.

METADATA

Both Email Classifier and Office Classifier add metadata to the email message or document. When a user sends or saves the content, the software stores machine readable copies of the labels chosen by the user. For email messages, labels are stored as x-headers. Most email servers, gateways and content scanners are able to apply rules based on the information stored in x-headers. For Office documents, metadata information is stored in the Custom Document Properties as defined by the Microsoft file format. Storing metadata in this location allows the labels chosen by the user to be included in the file even if there is no corresponding visual label (i.e. header/footer or watermark). Boldon James Email Classifier uses this metadata to perform Attachment Checking. Similarly, document management software can index the same properties to provide improved search results.

CLEARANCE CHECKING

Recipient validation is a process that can help to ensure that sensitive information is not mistakenly sent to a user who should not receive it. Boldon James Email Classifier has a feature called Clearance Checking that will verify that each recipient of an email message is allowed to receive the message based on the label chosen by the sender. Within the organisation, users can be assigned a Clearance that is stored as user object attributes within Active Directory (AD). As a user presses the Send button for an email message, Boldon James Email Classifier compares the label chosen by the sender for the message against the clearance information stored in AD for each recipient. If there is a user who has not been assigned a clearance level at least as high as the classification of the message, a dialogue box informs the user which recipients are not allowed to receive the message, and the message is not sent.

DATA CLASSIFICATION
DELIVERING VALUE

WHITE PAPER

EMAIL DOMAIN CLEARANCES

DOMAIN	CLEARANCE
*(Any email domain)	Public
MyOrg.co.uk	Public, Business-In-Confidence, Controlled, Board Members
MyOrg.com	Public, Business-In-Confidence, Controlled, Board Members
PartnerCompany.com	Public, Business-In-Confidence
MySolicitor.com	Public, Business-In-Confidence, Controlled, Board Members

In order to ensure that Clearance Checking can still be performed whilst a user is not connected to the organisation's network (i.e. Active Directory is not available), Boldon James Email Classifier adds clearance information to the Global Address List which can be downloaded to Outlook.

For recipients who are not part of the organisation's email domain, Boldon James Email Classifier has a policy tool for Domain Clearances. In this case, an organisation defines the clearance level for an email domain. For example, sending PUBLIC email messages is allowed to any domain, but sending CONFIDENTIAL information is allowed only to an approved set of email domains.

ATTACHMENT CHECKING

When a user sends an Office document as an email attachment, Boldon James Email Classifier can invoke a policy to verify that the attachments are not classified as having a higher sensitivity than the email message itself. For example, a user cannot send a RESTRICTED attachment within an UNCLASSIFIED email message. Boldon James Email Classifier can block the message and display a dialog box informing the user that there is a policy violation. The user can then either re-classify the email message or remove the attachment. For example, an email message labelled UNCLASSIFIED would be blocked if there was a RESTRICTED Word document attached.

MANAGEMENT

The configuration of Boldon James Email Classifier or Boldon James Office Classifier is performed with a single administration tool –with labels defined in common. Label definitions are stored in Active Directory and client settings are deployed via Group Policy.

The administration tool is also used to link labels to Microsoft's Rights Management Services (RMS). This feature allows users to choose a label from Boldon James Email Classifier or Boldon James Office Classifier and the software will automatically invoke RMS content protection. For example, a user may choose the sensitivity label BOARD MEMBERS, and the plug-in will encrypt the content using an RMS template allowing access to members of the board of directors.

The administration tool also provides simple interfaces for defining clearances and labelling rules. Clearance Checking templates can be defined for Active Directory groups or Organisational Units. For example, Boldon James Email Classifier can restrict recipients outside of the HR department from receiving email messages marked HR ONLY, based on membership of the associated Active Directory group. Labelling rules can be used to define required or optional labels and whether labels can be downgraded. For example, when a user forwards or replies to an email message, Email Classifier can restrict the label selection such that the new message cannot be of a lower sensitivity classification than the original message.

CONCLUSION

Information classification can benefit many different kinds of organisations, supporting security objectives as well as business process objectives. Benefits range from better user awareness about handling documents and email messages to a more efficient information management infrastructure. The challenges that such a classification programme introduces to both the users and IT departments are minimized by choosing a software tool that not only meets the requirements of information classification but is also user-friendly and easy for administrators to deploy.

DATA CLASSIFICATION
DELIVERING VALUE

WHITE PAPER

ABOUT BOLDON JAMES

For 30 years, Boldon James has been a leader in data classification and secure messaging solutions, helping organisations of all sizes manage sensitive information securely and in compliance with legislation and standards, in some of the most demanding messaging environments in the world.

Our Classifier product range extends the capabilities of Microsoft core infrastructure products to allow users to apply relevant visual & metadata labels (protective markings) to messages and documents in order to enforce information assurance policies, raise user awareness of information security and orchestrate multiple security technologies.

Our customers range from commercial businesses to Government, Defence & Intelligence organisations and we are a Microsoft Global Go-To-Market Partner and a Gold Application Development Partner. Boldon James is a wholly-owned subsidiary of QinetiQ, a FTSE 250 company, with offices in the UK, US, Australia and Europe and channel partners worldwide.

More Information

CLASSIFIER

FOR MORE INFORMATION ABOUT HOW YOU ADD BUSINESS VALUE AND ENHANCE YOUR ORGANISATION'S SECURITY USING BOLDON JAMES CLASSIFIER, PLEASE CONTACT US OR CALL +44 (0)1270 507800.

ALTERNATIVELY REGISTER FOR A LIVE DEMONSTRATION OF CLASSIFIER.