

# Boldon James

A QINETIQ company

## Classifier360: Balancing User Involvement With Control



### Including

- Classification methods
- Business drivers for classification
- The classification challenge
- Selecting the right approach
- The solution: Classifier360

## Introduction

Data classification enables organisations to add context to the unstructured information that they hold in messages, documents and files. This classification context allows the organisation to focus protection on the more sensitive data and ensure safe sharing whilst also gaining efficiencies in business processes and information management, thereby reducing cost.

There are various techniques available to implement data classification, each with their own nuances and benefits. These classification techniques are most commonly classed as either user-driven or automated.

Many organisations have found that incorporating user insight into the process of classification is vital. With user-driven data classification the organisation captures the user's knowledge of the context and business value of the data they create and handle, so that informed decisions can be taken about how it is managed, protected and shared.

Automated classification can ensure full coverage of classification is achieved across a variety of originating data sources – some of which are outside user control.

A successful data classification project will need to identify the right blend of classification techniques to deliver the greatest benefit, whilst meeting organisational requirements and ensuring users are effectively engaged. This paper explores how data classification can work in practice for organisations today, and explains how using the classification techniques provided by our Classifier360 system can ensure you achieve the optimal combination.

## The Classification Challenge

Once an organisation decides on implementing data classification, they first need to establish the schema or taxonomy by which to categorise information, together with the labelling or marking formats that users will be expected to recognise. The next step is to decide which classification techniques should be used to ensure the needs of the business are met.

It is now widely understood that using data classification software is essential to ensuring a consistent approach. Critically a software-based data classification solution will also add valuable metadata, which in turn can be used to direct the actions of complementary security solutions such as DLP, Rights Management or Encryption. Data classification solutions broadly fall into two distinct categories – automated classification and user-driven classification – and have their own distinct benefits.

### Automated Classification

Automated data classification has historically been provided by Data Governance or Content-Aware DLP solutions and uses software algorithms to analyse content in order to propose a classification for a file or message. These algorithms are commonly based on matching keywords or expressions found in the content.

An automated approach will normally be able to add metadata to record the result of its classification decisions. However, it is less likely that an automated system will be trusted enough to add the correct visual markings to the body of the document, as the document owner is not in a position to review the resulting changes to the content.

Whilst automated data classification offers an attractive range of algorithms that can be used to classify data, there is always an accuracy issue to contend with. Incorrect matches or “false positives” will always occur and the challenge is to tune these algorithms to provide an acceptable error rate that avoids frustrating users and business processes alike. Furthermore, the “false negatives” that occur when the system fails to identify sensitive data risks exposing the organisation to unnecessary data loss.

There are many examples of automated classification projects failing to deliver on original expectations as operational systems have to be detuned in order to strike an acceptable balance between user acceptability and accuracy.

The key challenges that are generally experienced when relying solely on an automated approach to classification can be summarised as follows:

- **Detection Errors** - errors in automated decision-making are inevitable, resulting in either search rules mis-categorising data (known as false positives) or failing to identify sensitive data (known as false negatives). Furthermore, some files simply cannot be categorised by textual analysis alone, such as determining which CAD files contain drawings with sensitive intellectual property (IP).
- **Low User Trust** - if users encounter a significant amount of misclassified data then this inevitably reduces their trust in an automated system and gives rise to considerable frustration if those errors impede business processes. For users to correct examples of misclassification they need to be able to distinguish between a system-applied classification and one applied by a user in order to assign the appropriate level of trust.

## User-Driven Classification

Many organisations already recognise the benefits of involving users in the data security process. User-driven data classification enables organisations to empower their user communities who create and handle data to assign value to it within a given context and in a language they understand. This contextual understanding is stored as visual & metadata labels applied to messages, documents and a wide range of files. A by-product of involving the user means that the organisation can realise further benefits including:

- Heightened user awareness of security policy and data value
- Increase in user trust as they are core to the decision making process
- Reduced number of classification errors
- Improved security performance as all users are actively involved in data protection
- Reduced business risk as data is more appropriately protected

For these reasons, organisations commonly find that involving the users is critical to the success of any data classification project. There are, however, some scenarios where relying exclusively on user-driven classification may not fully meet organisational requirements, for example:

- Business systems generate considerable volumes of unstructured data - such as reports from SAP or ERP systems
- When introducing classification in a phased manner, for example by first tackling large volumes of legacy and system-generated data before involving users in the classification process
- Where the organisation wishes to use automated techniques to guide users when classifying new material - here it can be important that a user can distinguish between a classification derived by the system, as opposed to one applied by a knowledgeable user, in order for them to understand what level of trust to place in the classification

## Selecting The Right Classification Approach

Organisations implementing data classification face the challenge of deciding which technique or blend of classification techniques to employ, as outlined previously.

In our experience these are some of the key considerations and common requirements organisations face when making this decision:

- You want users to classify material, but you want to support and guide them in their choices
- By offering intelligent defaults when classifying new material, based on a range of factors – such as the identity of the user, the origin of the data and attributes of the data itself
- You have data flowing into your organisation that needs pre-classifying to assist your users
- You have data generated by automated processes that should be classified at the point of creation without user intervention – for example reports that are produced by a ERP or SAP system

- You want users to be able to differentiate between classifications applied by automated processes and those applied by other users
- You want subject matter experts to have authority over specific classification decisions – for example for business critical or highly sensitive information (e.g. export control information)
- You want to start your classification project using automated classification but would like to engage users at a later stage

Capturing user insight within the process of data classification is important to ensure decisions around data protection are made within the correct context, something that a pure automated classification approach cannot alone deliver. However, there are scenarios where organisations may require some of the capability and functionality that an automated solution can offer, thereby supporting users in their decision making and to automate aspects of classification where manual user intervention is impractical or inefficient. For example ensuring system generated reports or files are appropriately classified before being distributed.

In summary, blending the use of automated techniques with user-applied classification will present a challenge unique to each business, but when successfully implemented it can deliver significant benefits for data safeguarding. In order to get this blend of techniques right for your business, your classification solution needs to offer an integrated range of approaches that you can tailor to meet your precise needs and that can be easily adapted as your needs evolve.

Classifier360 aims to solve some of the issues that arise from both user-driven and automated techniques into a holistic classification approach that comprehensively covers a wide variety of classification requirements.

## The Solution: Classifier360

Boldon James Classifier's primary focus on engaging users in the process of data classification forms the cornerstone of Classifier360, an Enterprise Classification System that blends together best practice in user-centric and automated

classification techniques in the manner most appropriate to your business.

The Classifier360 classification techniques are characterised as:

- **User-Driven Classification** - Users are empowered to make business-centric classification decisions
- **Recommended Classification** - Rules are used to propose a classification to the user, based upon attributes of the user and the data, such as the data type, metadata and content
- **Prescribed Classification** - Data is automatically classified without user involvement, either dynamically at the point of creation or in-transit, as well as retrospectively by using Discovery and Search tools
- **User-Endorsed Classification** - Supplements any of the other classification techniques by applying classification labels that require additional user endorsement before being regarded as authoritative

The manner in which you combine these techniques is driven by the needs of your business and its users.

Underpinning the successful delivery of Classifier360 is the Classifier Platform which offers unique integration and extensibility possibilities for Classifier customers and partners.

In addition, individual Classifier products augment the Classifier Platform to further expand the range of classification techniques available to your business.



## User-Driven Classification

Most of the Classifier products enhance the primary productivity tools and collaborative platforms used by your staff on a daily basis. As such these products directly support users in the process of classifying the data that they generate and manage. The user's insight and understanding of the true context and meaning of this data are accurately captured and utilised to ensure the data is correctly safeguarded.

That insight would normally be gathered via the user's decision as to the most appropriate classification. However, there are situations where the user can be assisted in that decision process by involving the rule-based and automated techniques of Recommended or Prescribed Classification.

## Recommended Classification

Recommended Classification provides an "intelligent" means of offering a default classification to the user where the material is otherwise unlabelled.

A range of contextual and content criteria can be used to offer a suitable default classification or to guide the user when classifying new material:

- **Identity** - or internal users their identity or membership can be utilised in assigning an initial classification, similarly the identity of external users can be used, for example when receiving external email
- **Data Type** - the type of data can be used in determining the classification, for example to distinguish an email from a document or file
- **Attributes & Metadata** - many data types will have attributes or metadata that can be used to derive a default classification
- **Keywords, Expression, Patterns, Algorithms** - the content of textual material can be used to guide the user in assigning an appropriate classification

## Prescribed Classification

Using Prescribed Classification, data is automatically classified without direct user involvement. This may take

place dynamically at the point of creation or in-transit, as well as retrospectively by using Discovery and Search tools or custom scripts and processes.

Classifier products and tools can be used to automatically classify data at the point of creation based on the following contexts:

- **Location** - for example, using a particular drive or folder location to determine the classification
- **Data Type** - for example, to assign a classification based upon the type of file
- **Attributes** - for example, to assign a classification based upon an attribute of the file
- **Source** - for example, to assign a file classification based upon the originating application

Similarly data in-transit can be automatically classified based on a variety of factors such as the source and destination context of a message.

Prescribed Classification techniques can also be used to automate the retrospective classification of existing data based upon either the selection rules of third-party Discovery and Search tools, or using the criteria of custom processes and scripts. For example, DLP Discovery agents may be used to locate data of interest and automatically apply classifications as part of their remediation actions.

## User-Endorsed Classification

User-Endorsed Classification is a technique unique to Boldon James Classifier and can be combined with any of the other techniques in order to defer final classification decisions to some or all of the user community in special cases. The technique is based upon creating a distinction within the labelling scheme between non-authoritative classifications, which indicate a "proposed" classification, and the authoritative classifications that can only be applied in specific cases by knowledgeable users as part of a defined workflow, for example.

By combining User-Endorsed with Prescribed Classification, a non-authoritative classification may be applied by an automated process such that a user may be guided

by that classification, but the user must then apply an authoritative classification in order to officially endorse the classification. This could be for very specific sensitive data, for example highly confidential information which requires a more authoritative decision. A scenario could be where an automated process applies the classification 'ASSUMED-CONFIDENTIAL' to a sensitive document such then when a user opens that document they must actively select a replacement classification such as 'CONFIDENTIAL'.

Similarly when combining User-Endorsed with Recommended Classification, rules are used to assign a default classification that is non-authoritative, requiring the user to then apply their authoritative classification choice – which may endorse or modify the original recommendation.

User-Endorsed Classification may even be combined with User-Driven Classification such that only nominated users may apply authoritative classifications to the non-authoritative selections made by other users. In this way it is possible to ensure that authoritative classifications relating to specific topics or information can only be applied by subject matter experts.

Where classification techniques are blended together, User-Endorsed Classification can be used to ensure that automated or rules-based techniques do not devalue the insight provided by users and that automated guidance never becomes submissive acceptance.

## Summary

With organisations struggling to protect the increasing amount of unstructured data, they need to employ a blend of data-centric protection techniques that caters for the changing security needs of their business. For organisations to meet the host of challenges outlined in this paper, the classification process should still be driven by users and informed by their knowledge of the business value of information. However, employing a range of classification techniques based on a well-understood and defined security framework can deliver significant business benefits from improving the performance of complementary security solutions, to increasing user awareness and transforming security culture.

To respond to the market's needs around a flexible and versatile classification solution, Boldon James has developed a comprehensive Enterprise Classification System in

Classifier360 which blends together best practice in user-centric and automated classification techniques in the manner most appropriate to your business.

At the core of Classifier360 is the importance of capturing user insight into the process of data classification. Users can define more accurate classification decisions as they understand context, something that can be missed by pure automated techniques. However it is also important to support users in their decision making and to automate aspects of classification where manual user intervention is impractical or inefficient. Blending the use of automated techniques with user-applied classification presents a challenge unique to each business, but when successfully implemented it can deliver significant benefits for data safeguarding.

Classifier360 builds upon the unique Classifier Platform to deliver a completely tailored solution for data classification across the Enterprise. The Classifier Platform offers unique integration and extensibility possibilities and augmented by specific Classifier products it provides Classifier360 with a broad spectrum of classification techniques to fit your precise business needs.

## Classifier360 Within Your Business

**Adapts** to your business and infrastructure needs

**Reflects** the differing requirements of your user communities

**Supports** users in their classification decision-making

**Streamlines** workflow for routine classification tasks

**Balances** technology-based decision-making with user insight

**Respects** the authority of user judgements **WIDENS** the reach of data classification

**Leverages** investment in Discovery tools such as DLP

## About Boldon James

Boldon James is an industry specialist, delivering globally-recognised innovation, service excellence and technology solutions that work.

Part of the QinetiQ group, a major UK plc and FTSE 250 company, we integrate with powerful data security and governance ecosystems to enable customers to effectively manage data, streamline operations and proactively respond to regulatory change.

We're a safe pair of hands, with a 30 year heritage of delivering for the world's leading commercial organisations, systems integrators, defence forces and governments.

- Pioneers of data classification and secure messaging solutions
- Technology that seamlessly integrates with data governance solutions
- Unrivalled customer service excellence and technical support
- Trusted by customers to reduce risk and streamline business performance

Global brands trust us to protect their sensitive data:



## More Information

For more information about how you can implement a data classification solution as part of your data protection strategy, please **contact us** or call **+44 (0)1270 507800**.