

# VARONIS WHITEPAPER

Five Steps to Faster Data Classification



# CONTENTS

UNSTRUCTURED DATA CHALLENGE	4
CLASSIFYING UNSTRUCTURED DATA	4
FASTER, MORE SUCCESSFUL DATA CLASSIFICATION	5
Identify Data Owners	5
Define Data of Interest	7
Use Meta-data to Focus and Accelerate	7
Report and Remediate	8
Rescan Data	9
CONCLUSIONS	9



# FIVE STEPS TO FASTER DATA CLASSIFICATION

Get actionable data classification results quickly by following the five steps outlined in this paper. We say “actionable” results because you’ll not only find sensitive data, but have the context required to remediate problems. Naturally, searching for sensitive data among your expanse of unstructured data is a process that requires automation – without automation, there’s simply no operationally efficient way to perform such tasks. The Varonis IDU Data Classification Framework is a software solution that’s optimized to address today’s large, growing and constantly changing collections of unstructured data, and makes performing these five key steps easy.



# UNSTRUCTURED DATA CHALLENGE

Unstructured data on shared file systems, NAS devices and SharePoint sites is a challenge to manage for most organizations. All of these spreadsheets, presentations, documents, multimedia files, etc., account for roughly 80 percent of business data, according to analyst firm IDC<sup>1</sup>. And, this shared data is highly dynamic: New data is constantly added, accounting for an average growth rate of 57 percent per year according to the same study by IDC. Some organizations see a doubling of unstructured data volume each year. The relevance of the data is also constantly in flux. Users may need access to data now, but not in a few months when their project finishes or when the data itself becomes outdated and stale.

## CLASSIFYING UNSTRUCTURED DATA

Organizations can quickly become overwhelmed with the task of managing and protecting this large, changing pool of data. As a result, more and more organizations are initiating data classification projects in the hopes of identifying their most sensitive data, remediating any problems, and implementing proper controls. Unfortunately, several challenges prevent data classification deployments from reaching their full potential. From a business perspective, a lack of actionable results is the primary challenge. Data classification solutions produce a list of files with sensitive content, but the question of what the files mean to the business and what to do with them is not inherently obvious.

From a technical perspective, the challenge is that data classification solutions scan every file looking for relevant content and are consequently slow to deliver results. And, on subsequent searches, these solutions must look at all files again, making it virtually impossible to keep pace with data growth and change.



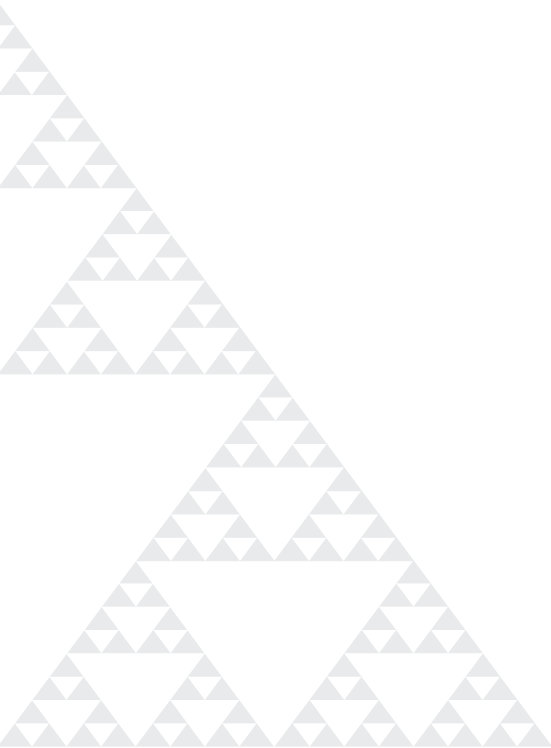
# FASTER, MORE SUCCESSFUL DATA CLASSIFICATION

The five steps below outline an approach for achieving data classification results dramatically faster than conventional methods. No matter what method you choose, you'll need software to help you automate the process of finding sensitive data – organizations simply have too much data changing too quickly to make manual processes viable.

## **1. IDENTIFY DATA OWNERS**

Data owners are at the heart of the process when it comes to managing unstructured data. Because they understand the importance of data assets to the business, data owners are a critical component to creating policies that make business sense. They can help determine who should and should not have access, what type of protections the data should have, and indicate when the data is no longer relevant to the business. When it comes to sensitive data, owners can decide whether data is at risk, and what

remediation steps are required. The problem is identifying and keeping track of data owners. A 2008 Ponemon Institute<sup>2</sup> study indicated that 84 percent of organizations were challenged to identify data owners. That's because the locations of data and the names of folders, directories or sites often provide few clues to true data ownership, and file system meta-data about data ownership goes stale quickly. Phone calls and emails, which are other common methods for identifying data owners, are simply not efficient or reliable enough to constitute viable processes either. Nevertheless, identifying data owners is a key part of the data classification process, and you will need an automated, repeatable way to identify and track data owners. One of the most effective ways to determine data owners is to track who is accessing the data. Over time, the top users of data will emerge, and these people will be able to tell you who owns the data. This is an area where Varonis DatAdvantage shines. Because it sees every data access made by every user, DatAdvantage can deliver the guidance you need to quickly identify the top users of data. Contacting one or two of these people, you'll either find the data owner, or these data users will quickly point you to the owner. And, with DatAdvantage, once you have identified owners for your data, you can keep track of who they are, and build them into your data classification and remediation processes.



# EXAMPLE FIVE-STEP APPROACH WITH VARONIS IDU DATA CLASSIFICATION FRAMEWORK

## **IDENTIFY DATA OWNERS**

Use Varonis DatAdvantage to identify and assign data owners for important data without a known owner. In this example, we'll say "Dave Smith" was determined to be the owner for the SharePoint site "Project X Resources."

## **DEFINE DATA OF INTEREST**

Work with Dave Smith to define what he considers sensitive. In this example, Dave says data containing the phrase "Project Budget" or the word "Secret" is sensitive.

## **USE META-DATA TO FOCUS AND ACCELERATE**

Dave says the sensitive data on his site should not be accessible to the Marketing or Operations teams. Use Varonis IDU Data Classification Framework to find data on the Project X Resources SharePoint site that is accessible by the Marketing or Operations teams, and contains sensitive content.

## **REPORT AND REMEDIATE**

Dave receives reports about sensitive data that is accessible by the Marketing and Operations team and works with IT Staff to change data permissions.

## **RESCAN DATA**

The Project X Resources site is rescanned weekly, and Dave gets reports showing whether any newly added or modified data violates his specified access policies.



## 2. DEFINE DATA OF INTEREST

Once you have identified data owners, you will need to work with them, as well as security and risk managers, to identify the key words, phrases and patterns that will identify the data of interest to your organization. Doing that successfully will require some investigative work on your part, and you'll need to understand what's driving the need to find data. For example, in many organizations, regulatory compliance is a driver. Regulations often specify which data is sensitive (e.g., personally identifiable information) and what measures are required to protect it (e.g., monitor access). Other common types of information requiring special attention are intellectual property, customer data and employee information. As you work with these parties, be as specific as possible about what needs to be identified and protected.

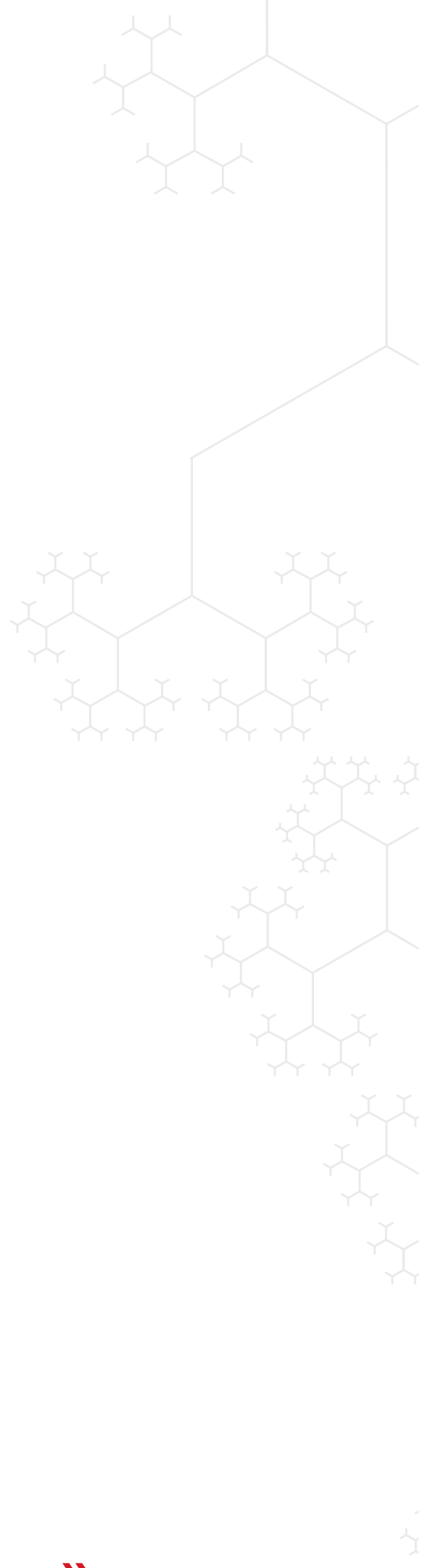
As part of defining what's of interest, it is recommended that you establish different levels of sensitivity based on the type of content your organization needs to manage and protect. Industry best practices show that a good rule of thumb is to constrain your hierarchy to four levels. With more than that, it becomes difficult and impractical to manage. Here are four example levels you can start with:

- “Secret data” is information that is critical to your organization’s core value. This is information that would be significantly damaging if, for example, competitors were to obtain it.
- “Confidential data”, or regulated data, is data that you are obligated to keep close control of due to regulations or to protect the privacy of your customers or employees.
- “Private data” is information your organization prefers remain out of the public domain, but is non-critical. An example might be the names of the product vendors your organization uses.
- Finally, “Public data” is data that is intended for public consumption, and is perfectly acceptable to share inside and outside the organization.

## 3. USE META-DATA TO FOCUS AND ACCELERATE

Meta-data is data about your data. Common examples are file sizes, types and locations, which all tell you about the data, but are not the data itself. When it comes to identifying sensitive data – and protecting it – a number of pieces of metadata are relevant and are discussed below. What is most important about using metadata, though, is that it can be used to focus and accelerate the data classification process. In this way, the meta-data becomes another element of the search. Specifically, it provides a short-list of where to look and what to expect.

For example, by first identifying those files that are poorly protected and then looking inside them for credit card data, you accelerate the process of identifying credit card data that is at-risk. Conveniently, organizations can use permissions meta-data to do just that. Any sensitive data found in overly accessible files has a clear remediation path: fix the access permissions to the data so that it is based on least-privilege (i.e., business need-to-know).



This is another area where Varonis DatAdvantage excels. The product builds a unique meta-data map of your business data and constantly updates the map with the latest information. In addition to collecting information such as permissions and data sizes, DatAdvantage also produces meta-data unavailable elsewhere.

The following are examples of meta-data that can be used to focus and accelerate data classification:

**Access permissions:** If you are trying to determine what data a specific user or group can (or cannot) access, permissions will tell you. And, a careful analysis of permissions can tell you which data is overly accessible. For example, data that is accessible by groups such as the Microsoft Active Directory “Everyone” group is clearly overly accessible.

**Access activity:** Data access activity provides important information such as which folders are the most frequently used and which folders are not being used at all. You can also determine which data was recently added or modified. That intelligence is tremendously useful, for example, in reducing the time spent searching. After the initial classification scan has occurred, subsequent searches can be restricted to just that data that needs to be classified (i.e., data that has not yet been searched). For specific users or groups, you can determine what data they have been accessing, which is important if you are looking to see who has actually been using the sensitive data.

**Ownership:** Ownership information helps limit searches to data owned by specific people. So, if you are working with individuals to help them get control over their sensitive data, this piece of meta-data will narrow sensitive data searches to just the relevant data.

#### **4. REPORT AND REMEDIATE**

Generating results is obviously an important part of classification projects, but it’s not the final stage. Once you have results, you’ll need to get these into the hands of decision-makers – which are typically data owners and governance/risk/compliance teams – so that these people can understand the situation and begin formulating remediation strategies and plans.

Varonis DatAdvantage couples its ability to designate specific individuals as data owners – mentioned above – with enterprise-class reporting to produce special reports for data owners that provide them with constantly updated business intelligence about their data.

Data owners are important stakeholders for results because they are typically in the best position to identify exactly what the content is, whether the data is stored in the right place, and who should and should not have access to it. Data owners can also help build a remediation strategy and process, especially once they are armed with specific examples involving their own data. Governance, risk and compliance staff can provide the overall oversight needed to ensure data is being protected in accordance with the organization’s objectives. And, these teams can use result reports as the basis of documentation for audit requirements.



## 5. RESCAN DATA

Recall, from the beginning portion of this article, that unstructured data is growing at about 60 percent annually and is constantly being modified and going stale over time. Based on those facts, there is a need to periodically re-scan data to ensure that you maintain an accurate view of your sensitive data. Ideally, you would limit your searches to newly-added data to determine if it contains sensitive information and to existing data that has been modified to determine if it has either gained or lost relevance to your classification project. This is another area where Varonis DatAdvantage benefits from meta-data use. It is able to accelerate the rescanning process by looking only at newly-added or modified data, slashing rescanning times. Finally, you'll want to provide data owners and governance, risk and compliance staff with updated intelligence based on your re-scanning.

# CONCLUSIONS

Searching for sensitive data in your unstructured data stores requires the use of data classification software: this data is simply too voluminous and dynamic to process and manage manually. A solution that allows you to implement the five steps outlined above – especially incorporating meta-data into the process – is critical for achieving actionable results. Without meta-data, data classification projects can take far too long, and the results they produce typically don't have the context required to remediate problems. Using meta-data dramatically cuts the time it takes to produce results and provides the context required for problem remediation.

<sup>1</sup>IDC, *"The Expanding Digital Universe"*, 2007

<sup>2</sup>Ponemon Institute, *"Survey on the Governance of Unstructured Data"*, 2008

# ABOUT VARONIS

Varonis is the leading provider of software solutions for unstructured, human-generated enterprise data. Varonis provides an innovative software platform that allows enterprises to map, analyze, manage and migrate their unstructured data. Varonis specializes in human-generated data, a type of unstructured data that includes an enterprise's spreadsheets, word processing documents, presentations, audio files, video files, emails, text messages and any other data created by employees. This data often contains an enterprise's financial information, product plans, strategic initiatives, intellectual property and numerous other forms of vital information. IT and business personnel deploy Varonis software for a variety of use cases, including data governance, data security, archiving, file synchronization, enhanced mobile data accessibility and information collaboration.

## Free 30-day assessment:

### **WITHIN HOURS OF INSTALLATION**

You can instantly conduct a permissions audit: File and folder access permissions and how those map to specific users and groups. You can even generate reports.

### **WITHIN A DAY OF INSTALLATION**

Varonis DatAdvantage will begin to show you which users are accessing the data, and how.

### **WITHIN 3 WEEKS OF INSTALLATION**

Varonis DatAdvantage will actually make highly reliable recommendations about how to limit access to files and folders to just those users who need it for their jobs.

#### **WORLDWIDE HEADQUARTERS**

1250 Broadway, 31st Floor, New York, NY 10001 **T** 877 292 8767 **E** [sales@varonis.com](mailto:sales@varonis.com) **W** [www.varonis.com](http://www.varonis.com)

#### **UNITED KINGDOM AND IRELAND**

Varonis UK Ltd., Warnford Court, 29 Throgmorton Street, London, UK EC2N 2AT **T** +44 0207 947 4160 **E** [sales-uk@varonis.com](mailto:sales-uk@varonis.com) **W** [www.varonis.com](http://www.varonis.com)

#### **WESTERN EUROPE**

Varonis France SAS 4, rue Villaret de Joyeuse, 75017 Paris, France **T** +33 184 88 56 00 **E** [sales-france@varonis.com](mailto:sales-france@varonis.com) **W** [sites.varonis.com/fr](http://sites.varonis.com/fr)

#### **GERMANY, AUSTRIA AND SWITZERLAND**

Varonis Deutschland GmbH, Welschstrasse 88, 90489 Nürnberg **T** +49 (0) 911 8937 1111 **E** [sales-germany@varonis.com](mailto:sales-germany@varonis.com) **W** [sites.varonis.com/de](http://sites.varonis.com/de)