

An abstract digital background with a dark blue and purple color palette. It features a central bright light source that creates a lens flare effect, with numerous thin, glowing lines radiating outwards, suggesting a data network or digital space.

DARK DATA, BIG DATA, YOUR DATA:
CREATING AN ACTION PLAN FOR INFORMATION GOVERNANCE

A WHITE PAPER BY VIEWPOINTE®

TABLE OF CONTENTS

- Executive Summary3
- Dark Data: Addressing the Unmanaged Content at the Edge of Your Governance Program3
- Dark Data Can Hold Hidden Risks.....4
- Big Data and Big Content Needs Big Governance4
- Take Action on Dark Data: Defensibly Dispose or Light Up the Value.....5
- Four Elements of a Dark Data Action Plan.....6
 - Analysis.....6
 - Classification7
 - Collection7
 - Assessment8
- Stay Enlightened: Avoid Dark Data with Modern Approaches to Information Governance.....9
- Viewpoint: A Trusted Partner for Big Data and Information Governance.....10
 - Information Governance in an Era of Big Data and Cloud10
 - Domain Expertise and Content Assessment Services.....10
- Conclusion11

EXECUTIVE SUMMARY

Unmanaged, uncategorized, unknown content is widespread in most enterprises. This neglected information occupies valuable storage capacity and may contain hidden risks and typically lies abandoned at the periphery of your information governance program. “Big Data” projects that support customer service, marketing or process improvement initiatives can open the door to even more dramatic accumulation of data that quickly goes dormant. Described as “dark data,” this unused or unidentified content sits outside of the retention schedules, classification schemes and retrieval systems that organizations rely upon to meet compliance obligations, ease the burden of electronic discovery and ensure decisions are made with accurate and relevant information.

As the volume of enterprise information continues to rise at unprecedented rates, information professionals, risk managers, IT teams and business peers must work collaboratively and take action on dark data. A dark data action plan can serve as a roadmap for organizations committed to dealing with the unmanaged content sitting at the edges of the enterprise. Four essential activities – analysis, classification, collection and assessment – can help your organization make informed decisions about dark data.

Technology, industry best practices and new generation governance platforms can help large enterprises gain insight into which dark data can be illuminated and leveraged to provide business value, and which should be defensibly deleted in line with corporate policies.

DARK DATA: ADDRESSING THE UNMANAGED CONTENT AT THE EDGE OF YOUR GOVERNANCE PROGRAM

Peering into the shadows of your business’s information storage closets can be a daunting task for some information professionals. What unknown content resides on those cobwebbed disks and last century’s archive servers? Forgotten copies of spreadsheets? Email inboxes of long-departed staff? Decommissioned financial systems?

Storage demands are escalating and the cost of finding information when the lawyers and

auditors get involved can be astronomical. Regulated industries may be at risk of fines or sanctions due to content that lies unattended and unprotected on old storage disks, keys or tapes. Consistent execution on an information governance strategy is difficult when so much corporate content sits “off the radar” of IT, compliance and legal teams.

The term “dark data” has surfaced to describe this category of unmanaged content.¹

DARK DATA CAN HOLD HIDDEN RISKS

Information professionals in IT, records or compliance roles need to become aware of the information that lies at the periphery of their governance programs. Information at the edge of enterprise awareness can be a drag on productivity, infrastructure efficiency, search and discovery. This unmanaged, forgotten data may even conceal obsolete or inaccurate information that could be misinterpreted if discovered by auditors or attorneys.

All forms of electronically stored information (ESI) may be subject to legal discovery if a threat of litigation emerges – even obsolete or incomplete data. The presence of uncategorized, unmanaged dark data can result in increased costs of the find, review and analyze phases of

discovery. Increased risks may also result if this dark data includes unidentified drafts or duplicates of documents that should have been disposed of in line with retention policies.

Information professionals are starting to consider how technology and fresh approaches to information governance can put this questionable content back on the corporate data map. The Sedona Conference, a noted nonprofit research institute for the advanced study of law, recommends that organizations “reduce the ongoing accumulation of inactive information”² and that “an organization should avoid excessive retention of inactive information by destroying it when it is no longer necessary to meet legal retention requirements or business needs”.³

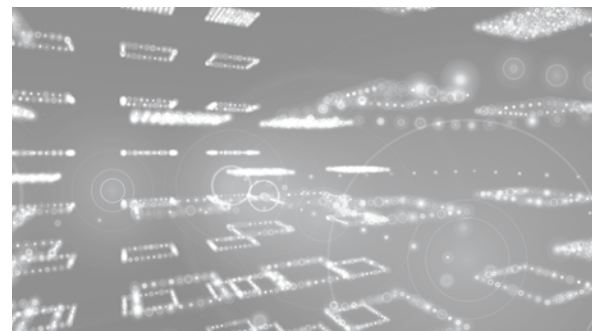
BIG DATA AND BIG CONTENT NEEDS BIG GOVERNANCE

The terabytes of dark data inside the enterprise are at risk of expanding exponentially as more and more organizations explore big data and social media initiatives. Enterprises have been handling large sets of data for decades, but recently “Big Data” has become a buzzword. Many organizations, including regulated businesses are actively planning big data initiatives.⁴

The definition of big data continues to evolve. Descriptions such as “volume, velocity and variety”⁵ and the “frontier of a firm’s ability to store, process and analyze”⁶ information are emerging from analyst firms. Big data reflects not only how an organization identifies, analyzes and uses the data managed within its own walls, but also data that was previously considered inaccessible, including data from new sources of information that may lie outside the control of an organization, to make business decisions. It’s

about making sense of information that – until recently – was too expensive, too time consuming or too difficult to access.

Big data includes unstructured forms of content that may be new to your information governance radar: rich media, web content, audio and messaging. Enterprise applications, such as ERP, HR and CRM systems, are also now incorporating more social and mobile



capabilities into their core platforms that add to the amount of information that can be mined. This rapid growth of electronically stored communications adds to the already large storage footprint of entrenched business systems and the resulting backups, archives, log files or export sets that must be managed to avoid adding to the data debris pile of redundant, outdated and trivial content (ROT) that impacts efficiency and adds undue risk.

In the era of Bring Your Own Device (BYOD), the new sources of information that lie at the edge of your enterprise add mobile, web and socially generated information such as sensor data, rich media or transactional log files that are used to make decisions. Once utilized, big data content, previously unavailable to your employees, may be relegated to unmanaged storage media, putting this information on the periphery of a governance plan. The value of big data is quickly extracted; however, it can quickly add to the volume of dark data over time.

The rise of the “internet of things”⁷ – the increased use of remote assets to provide and integrate data into existing business processes – enables more accurate decision support. This now means that big data can include structured data from mobile and sensor devices. In fact, IDC research in 2012 identified machine-generated data as the fastest growing source of information.⁸ This transactional information, including ESI for the purposes of eDiscovery, may be readable only with specialized analytic or business intelligence tools. The risk of this data “turning dark” after the initial real-time or near-time insights have been extracted can be high. The volume, variety and velocity of these forms of big data can mean risk if extracts, copies or exported sets are not managed.

Appropriate management might mean retaining insights and reports with business or legal value, and ensuring the residual data is also safely scheduled for disposition.

TAKE ACTION ON DARK DATA: DEFENSIBLY DISPOSE OR LIGHT UP THE VALUE

Assessing the dark data in your enterprise can be an onerous task if tackled manually. Few organizations have the personnel resources to search through the depths of unmanaged storage environments. The magnitude and scope of such a task is overwhelming and prevents many organizations from taking decisive action. Failure to act, however, will inevitably exacerbate the situation. Each day – often 24/7 – new data is created, transactions are captured and new sources of content are adopted by customers and business management.

You can choose to assert control over your dark data with a plan, the right tools and a

methodology designed to shed light on the unknown. The benefits of taking action should be viewed through the lens of economics, compliance or productivity.

- **Economic benefits** can include reduced storage costs by eliminating redundant services or devices. Organizations embarking on an infrastructure modernization initiative can begin highlighting areas of savings by discontinuing expensive shelfware – software that sits underused but incurs annual maintenance fees – and by redirecting hardware upgrade costs into modern, cloud-based services.

DARK DATA, BIG DATA, YOUR DATA: CREATING AN ACTION PLAN FOR INFORMATION GOVERNANCE

- **Compliance benefits** may include a lowered risk of policy breach by keeping obsolete information beyond its scheduled retention period, or reduced risk of fines by storing regulated information such as personally identifiable information (PII) outside of appropriately-secured systems. The risk of sanctions that result from incomplete or inaccurate disclosure during audit or eDiscovery may also be reduced.
- **Productivity benefits** can be achieved by removing potentially confusing, outdated or duplicate sources of information from the fingertips of your busy information workers. Time is wasted daily sifting through duplicate files, figuring out correct versions and

Assessing the dark data in your enterprise can be an onerous task if tackled manually.

reconciling contradictory data. Productivity gains can also be realized when data that was formerly left dark and unused is reclaimed, categorized and put to use. Records of past projects or historical trend data can be important, particularly in organizations committed to using analytics for performance measurements or decision-making.

FOUR ELEMENTS OF A DARK DATA ACTION PLAN

Your dark data action plan should include four key tasks: analysis, classification, collection and assessment. The outcome of these four activities is a roadmap for addressing the dark

data building up at the edges of your enterprise. The roadmap will help you make informed decisions about what to do with dark data – what to light up and what to defensibly delete.

1. ANALYSIS

The first phase of a dark data action plan is to understand your uncategorized, unorganized, unknown content. Analysis helps reveal what the data is, the format it is in, whether it is a duplicate and how much storage capacity it occupies. The goal of the analysis phase is to identify what's valuable and what is considered ROT.

Research conducted in 2011 found that 74% of enterprises still rely on manual techniques for categorization of information.⁹ However for most large enterprises attempting to do an analysis manually is a recipe that often fails. For effective results from an assessment

exercise, leverage technology to tackle the volume of inactive information lying dormant in your enterprise. Advanced search and content analytic tools have evolved to a level of maturity that is, at last, approachable for most businesses. Sophisticated analysis and extraction of concepts, entities and metadata are wrapped into user interfaces and graphical views that put the power in the hands of information managers rather than just specialized developers. Licensing and pricing models have also become more approachable as the rise of cloud- and subscription-based tools has lowered the barriers to entry and overall total cost of ownership.

2. CLASSIFICATION

As human efforts alone are typically costly, time-consuming and likely subject to high error rates, automation of the classification and categorization process is essential when tackling large volumes of dark data. A sustainable, realistic action plan will make use of technology wherever feasible.

Content analytics with auto-classification capabilities can be a key element at this stage of your dark data assessment project. Unknown, uncategorized information sitting in the dark data pool can be opened for analysis with software. Content analytic tools can match unstructured information – such as documents – against an existing taxonomy. Clusters of content can be grouped and recommended categories identified, aiding classification at a large scale. Category schemas can be self-trained as the classification processes are run, tuning results and creating more accurate, enforceable results. Outliers can be identified, revealing the types of content that fall

outside your existing taxonomy, helping to highlight gaps in the schema. Focus the expertise of your information managers on exception-handling – not routine tasks. Let technology perform the high-volume, predictable analysis work.

Classification is an important step in separating the dark data worthy of illumination from the redundant, obsolete and trivial data. Classification helps identify the ROT pile for safe and defensible disposition. Organizations responsible for protected information can benefit from analytical tools in order to uncover hidden pockets of data that can present risk if not properly managed. Auto-classification technologies can be configured to find and flag content that contains PII and regulated data such as credit card numbers, monetary amounts or government-issued ID numbers. Pulling this protected information into an appropriate classification scheme or applying approved retention schedules will be an essential part of closing any compliance gaps.

3. COLLECTION

The third element of your dark data action plan is to make decisions about ongoing management and governance of the information. The classification phase of the plan should have significantly whittled down the volume of data worthy of deeper analysis and protection. The ROT has now been scheduled for disposition in accordance with approved retention policies.

Collection helps to shed light upon the remainder of this once-neglected data, helping to ensure that any residual business, regulatory or historical value is extracted, managed and made available to relevant users.

Ongoing management and governance of this newly organized information can be streamlined with technology. An information governance platform with secure repository capabilities, retention and disposition management tools,



search, discovery and retrieval functions should be considered, to help ensure information is available to information managers, compliance and legal officers and business users. Secure private cloud platforms provide flexible, on-demand scaling and enable organizations to modernize their infrastructure. Sensitive information can

be locked down, records required for legal or regulatory compliance can be preserved, and content that contributes to corporate knowledge can be shared with peers. Retention rules can now be applied, bringing consistent lifecycle management to this recovered data, and helping to ensure it is disposed of when appropriate.

4. ASSESSMENT

The final phase of a dark data action plan is to perform a thorough review of your information landscape. An assessment helps determine important details about your information governance landscape. Where is your content being held? What applications, repositories or storage platforms are in use? What kinds of documents are being captured and stored? Are new forms of documents being adequately addressed as mobile, social and web applications rise as sources of content? How closely are your retention and disposition rules being followed? Are processes consistent, defensible and designed to support policies and regulations? Is confidential or protected information, such as PII or client data, appropriately secured and monitored?

Now that all of the previously unmanaged, unknown information has been located, organized, preserved or disposed of in

line with retention policies, it is time to document your current state and plan for the future. Large, complex enterprises will often seek advice from external sources such as industry analysts, information governance experts, solution providers or integrators. External perspectives on benchmarks, best practices and methodologies can help ensure your information governance roadmap is built on a foundation of knowledge and expertise.

Recommended activities when performing an assessment include:

- Evaluating sample sets of data using auto-classification tools, in order to identify any gaps or deficiencies in taxonomy
- Highlighting areas of risk for remediation or proactive mitigation
- Identifying cost-reduction opportunities, such as retirement of unneeded storage capacity, decommissioning of obsolete applications or elimination of manual processes, reduction of physical storage space
- Benchmarking against industry-accepted maturity models or frameworks to better understand your organization's information governance readiness and prioritize any required improvements

**An assessment helps
determine important details
about your information
governance landscape.**

STAY ENLIGHTENED: AVOID DARK DATA WITH MODERN APPROACHES TO INFORMATION GOVERNANCE

Businesses are eager to leverage technology to solve their problems faster, cost-effectively and by avoiding new hardware lock-in commitments. Research conducted in 2012 by the Information Systems Audit and Control Association (ISACA) and the Cloud Security Alliance revealed that the desire to experiment with new platforms and the need to find new ways of enhancing efficiency and employee effectiveness were among the top drivers for businesses adopting cloud, SaaS or consumption-based approaches of technology. Organizations committed to modernizing their IT infrastructure and embarking on a path to information governance across the enterprise can leverage these new tools, approaches and cost models.

Valuable resources – people, time and money – can be invested in delivering services and meeting objectives. No longer must scarce program dollars be absorbed by never-ending maintenance of underused software licenses or near-capacity hardware devices.

Elimination of the dark data hiding in the digital cracks and corners of your enterprise is an important milestone when establishing an information governance strategy that can withstand scrutiny by auditors, attorneys and compliance officers. But it is not the only step. Ongoing commitment to a sound information

governance strategy is not an easy task, but it is achievable when built upon a strong foundation of technology, consistent policies and processes, and employee awareness. Technologies such as content analytics, search and auto-classification have evolved in recent years, and new solution providers are making these technologies accessible, approachable and usable by offering consumption-based service models. The upfront capital and license cost burden of traditional on-premises solutions are becoming obsolete propositions, allowing enterprises to try new, more agile approaches to critical problems.

Look at the information landscape of your enterprise. Data maps, content inventories, taxonomies or classification schemes reveal a lot about the sources, locations and uses of your business information. A sustainable approach to information governance will ensure that the tools you use are flexible enough to scale with increased volume, are modular enough to handle content from a broad range of originating applications and are well suited to all end-users to promote adoption. Email servers, messaging platforms, SharePoint sites, enterprise applications, content management systems and/or social media platforms: you will need to address all content silos to help ensure that your data doesn't turn dark once again.



VIEWPOINTE: A TRUSTED PARTNER FOR BIG DATA AND INFORMATION GOVERNANCE

Information governance is more than content or data management, more than records and retention schedules, more than mere storage. Information governance is a strategic approach to the management of all enterprise information – from creation to final disposition – and helps organizations control costs and eliminate the risks presented by too much ROT to compete more effectively. Taking action on the dark data gathering inside your business is an essential part of a governance program.

For more than a decade, Viewpointe has successfully helped large, complex

organizations manage, secure and monitor their data via our hosted digital archive. With over 30 petabytes of data under management, Viewpointe offers the scale, security and best in breed infrastructure needed to meet the most complex information management and governance requirements to increase efficiencies, mitigate risk and reduce cost. Viewpointe's information governance platform, OnPointe®, and our team of experts can help you conquer your information management initiatives, including building a dark data plan that aligns with your information governance strategy.

INFORMATION GOVERNANCE IN AN ERA OF BIG DATA AND CLOUD

The OnPointe platform is the next generation of Viewpointe's secure, private cloud archiving platform that has served large financial services institutions for more than a decade. OnPointe is an integrated, holistic approach to the assessment, archiving, management, use, preservation and safe disposition of your business information. Good information governance means consistent security, classification and retention policies for digital content regardless of where it resides in its business lifecycle. OnPointe delivers on

the promise of flexible, agile cloud services without sacrificing the security and compliance demands of regulated business.

Five key capabilities comprise the OnPointe information governance platform:

- Content and data archiving
- Search, retrieval and discovery
- Security and privacy controls
- Retention management
- Secure private cloud services

DOMAIN EXPERTISE AND CONTENT ASSESSMENT SERVICES

OnPointe adoption is enhanced with expert services to guide content assessment, implementation and maintenance. We're ready to help build your dark data action plan and information governance strategy today. Service-level agreements provide trusted, managed

services designed to meet the needs of many types of complex, regulated businesses. The OnPointe information governance platform consists of proven technologies, integrated and delivered in a secure private cloud. Designed to be extended with modular

solutions that support major sources of enterprise information, the platform offers a consumption-based pricing model. This model means paying for only those managed services needed to execute on your information governance needs. The OnPointe approach to information governance means no costly shelfware, flexible scaling as the need for capacity grows and fresh opportunities to eliminate expensive hardware capital investment.

Viewpointe's cloud-based information governance platform helps you gain control and manage the lifecycle of your business information to realize the following benefits:

- **Analyze** unmanaged enterprise data (dark data) via Viewpointe's content assessment program to help you decide what information to manage, leverage and trust based on value
- **Manage** the explosion of data by archiving content based on value while defensibly disposing of ROT



- **Enforce** corporate retention, disposition and legal hold policies consistently across the enterprise to help ensure compliance and mitigate risk
- **Reduce** the burden on end users and improve your ability to collect and classify information through automated and consistent processes
- **Respond** quickly and cost effectively to eDiscovery, audit and internal investigation requests while increasing predictability of cost
- **Control** costs for lower total cost of ownership (TCO) through minimized capital investments and IT overhead via OnPointe, Viewpointe's private cloud solution

CONCLUSION

Dark data has likely been accumulating inside your enterprise for years and is now growing at an accelerated rate. The burden of storage and migration costs of this unused, uncategorized content is, perhaps, outweighed only by the unknown level of risk that could be lurking within it. Outdated, inaccurate, out-of-context information could have negative consequences, including fines or sanctions, if found as part of an electronic discovery order or audit.

Some dark data contains useful business information and is worth shedding light on

for users to leverage. Lost context, background information or trend data can help inform business decisions. Creating and executing a dark data action plan is a valuable first step in addressing this challenge.

The rise of secure cloud computing, consumption-based pricing and maturing approaches to auto-classification, retention management, archiving and search means that even large organizations with daunting dark data challenges have viable options to help achieve your information governance objectives.

DARK DATA, BIG DATA, YOUR DATA: CREATING AN ACTION PLAN FOR INFORMATION GOVERNANCE

END NOTES

- 1 Gartner Inc. blog. "Dark Data is like that Furniture You Have in that Dark Cupboard," 2012.
- 2 The Sedona Conference. "Commentary on Inactive Information Sources," 2009. See page 2.
- 3 The Sedona Conference. "Commentary on Inactive Information Sources," 2009. See page 2.
- 4 IBM. Big Data Survey highlighted in TechWeekEurope, 2012.
- 5 Gartner Inc. blog. Definition of Big Data, 2012.
- 6 Forrester Research. Forrester Definition of Big Data, 2012.
- 7 "The Internet of Things is Moving to the Mainstream," Gartner, January 10, 2013.
- 8 IDC Research Predictions for 2012, 2012.
- 9 Kahn Consulting. "How Do You Scale an Information Mount Everest?" 2012. See page 4 for data from the 2011 survey conducted by the Council for Information Auto Classification.
- 10 ISACA. "2012 Cloud Computing Market Maturity Study Results," 2012. See page 15.



Viewpointe®, a leading cloud provider of information governance and payments exchange services, is recognized for its national, trusted archive with more than 30 petabytes of data under management. With expertise in serving highly regulated financial institutions, Viewpointe offers the scale, security and state-of-the-art infrastructure needed to meet the most complex content management and information governance requirements. Viewpointe's information governance platform, OnPointe®, leverages your existing technology investments and streamlines your information governance processes, management, access and storage via a private cloud, helping you mitigate risk, increase efficiencies and reduce costs.

If you would like to learn more about how OnPointe can help optimize your information governance, management and storage environment, contact an OnPointe Advisor today.

(800) 956-3807 | onpointeadvisors@viewpointe.com | viewpointe.com/onpointe