CRAWFORD TECHNOLOGIES

WHITE PAPER

Shedding Light on Dark Data

Extending the Value of Document Archives

wwww.crawfordtech.com sales@crawfordtech.com

Shedding Light on Dark Data Contained in Document Archives

Transactional documents are the foundation upon which business process is built. Without things like invoices, statements, bills and business reports, workflow in most organizations would simply grind to a halt. The data contained in these documents facilitates the process and the "documentation" verifies the actions taken and the service provided. Indeed, choose any process and you can bet that there is at least one document, if not several, that are essential to the job.

New content management tools help organizations access large amounts of data stored in customer communications archives and reuse in important analytics applications

As a result, organizations in all industries have for years saved and stored transactional documents, often in huge volumes, in various archives and repositories. While this approach works well to provide a repository of proprietary print files and document images, it does not allow organizations to capitalize on the valuable data locked within these documents by easily extracting and analyzing it. Specific points of information like amounts, items, dates and account numbers are essentially "dark data" – unseen and often forgotten in terms of its accessibility and usability for business intelligence and process improvement. Plus, companies find that when a demand for legal discovery arises, or it's time to comply with an audit, this dark data can represent a very real risk to the organization. And extracting it can be time-consuming and expensive, if not impossible, to achieve.

What is needed are next-generation content management tools that enable organizations to access large amounts of data stored in customer communications archives and prepare it for reuse in important analytics applications. Indeed, the ability to retrieve and analyze customer and business data spanning years and decades of activity provides a more complete picture and is a powerful tool to boost business intelligence and improve information governance overall. This previously overlooked or unseen data can yield significant operational advantages, especially when it is combined with structured data for analysis, decision-making and discovery.

The opening up and leveraging of transactional document archives changes the traditional business equation by transforming these archives from a mandatory cost and risk obligation to a source of business intelligence which can lead to operational and competitive advantage.

Why is all of this dark data important?

Business Intelligence – using data for competitive advantage

Many C-level executives wish they had a crystal ball, but the closest thing available may be thought of as "business intelligence." The term dates back to the 1860s when it was coined to describe how businesses gained profit by understanding and acting upon information about the market prior to competitors. Today the notion has re-emerged to become an umbrella term to describe modern business decision-making using digital fact-based support systems. We all know there is no crystal ball but, the best way to predict the future is to fully understand the past.

Imagine the ability for knowledge workers to specify a set of historical documents from which data will be extracted from thousands or millions of invoices, statements, bills, reports and other documents. That is what business intelligence is all about - acting upon information to gain an edge on your competitors and in the market. But that is difficult to do if important business information is locked inside your transactional document archives. As a result, organizations today are increasingly looking for ways to shed light on this dark data and use it for more thoughtful business intelligence and strategic planning.

One example is a major credit card company that wants to extract data from their customer statements originating from a specific region, over a specific time period, and for only one of their many lines of business. How would you achieve that today? Or consider a healthcare insurance provider that each day publishes hundreds of thousands of explanation of benefits. Each customer communication carries data like a diagnosis codes, patient record numbers, addresses and amounts. How would you investigate a specific diagnosis code related to a specific zip code? The undertaking may indeed be so vast and complex that the project is never completed and, as a result, have has a direct (and negative) influence on the design of regional coverage options and patient wellness programs.

C-level executives wish they had a crystal ball, but the closest thing available is "business intelligence"

In both examples, without next-generation tools such as CrawfordTech's Dark Data Feeder, the analysis would take months to complete and require hundreds of staffing hours. Other examples are found in all industries and market segments; including financial services, manufacturing, government operations and more. In the increasingly competitive business climate, organizations can no longer afford to live without the business intelligence and operational advantages which are lying untapped in the dark data of transactional document and customer communication archives.

Discovery – dynamic document building

A court ordered discovery demand for information is another good example of how dark data in document repositories can represent a significant risk and impediment to organizational performance. Dynamically extracting documents and building those documents is imperative, but as the volume and variety of information in e-discovery cases continues to skyrocket, companies feel the pinch in terms of cost, risk and effort.

Many content management systems excel at managing, storing and outputting information. The problem is not with document output - it is with extracting the correct documents and assembling them into a single format even when the documents originate in disparate repositories and are store in multiple disparate formats. You can search and view, but most systems do not provide efficient ways of dynamically building documents from disparate sources so that the information can be easily presented and understood. This means that for any significant discovery demand, teams of analysts and clerks are required to assemble each result one document at a time. This takes time, additional technical support, and significant expense in staffing and labor.

Consider that over 19 million civil cases are filed in United States courts each year. The annual cost for these civil lawsuits is between $200 and $250 billion with up to 50 percent incurred to perform discovery. And that does not include soft costs like business interruption and stock market impacts. Clearly, organizations from all industries today are looking for ways to speed discovery efforts and cast a wider net when it comes to dark data and legal discovery.

The good news is that modern software (like our own Riptide), can perform a federated search across multiple archives, extract disparate document types, convert them all into PDF, and then assemble the content for submission to a court or other requesting authority.

Regulatory – mitigate risk, maximizing compliance

One of the reasons for having a transactional print and customer communications archive in the first place is the need to adhere to increasingly strict regulatory requirements. Scores of different regulations and standards dictate which documents must be kept, in what format, and for how long. Seven years is a common timeframe in many industries, but others like manufacturing, automotive and pharmaceutical require longer periods of time spanning ten, twenty five, or more years. Some documents, like civil engineering documentation, for example, must be kept indefinitely. And when no specific retention requirements exist, most organizations choose to err on the side of caution and keep duplicate and redundant copies – often printed out on paper – to ensure that the organization will withstand unanticipated regulatory and legal scrutiny.

Meaningful extraction and analysis of dark data housed in document stores requires advanced solutions and strategies that drive the value to the organization.

Document retention requirements can vary and are influenced by region or location as well. Statutes of limitations differ from state to state and the potential risk of a claim varies greatly from industry to industry. For example, in the United States, each state law differs as to what must be retained. This is a particular challenge for nonprofits and government operations. The question becomes: what types of documents are within the scope of the requirement? In some cases the length of time to retain a document is governed by the time period that a potential claimant has to bring a claim in that state.

Important Considerations

Multiple Archives

It is common for enterprise organizations to have multiple document archives and repositories of information. For example, one for transactional print data, another for marketing and customer correspondence, and others for line-of-business activities like claims processing, contract management or product development. Rarely do these systems work in concert with each other; especially when it comes to extracting documents and/or analyzing data across these diverse repositories.

Meaningful extraction of documents and analysis of data housed in multiple document stores requires advanced solutions (such as Riptide or Dark Data Feeder) as well as more comprehensive strategies that drive the value to the organization

Document Formats

Another important success factor is the ability to deliver the information contained within the documents in a format consistent with what the requester wants. For example, many transactional document archives contain information stored in AFP format. “Advanced Function Presentment” is the electronic architecture and print protocol originally developed by IBM that is now a de-facto standard for high-volume production printers. This is an important consideration because regulatory requirements often stipulate that customer information be stored in the way it was originally produced. And while an original production AFP file is great for printing, it’s not great for electronic presentation. It is important therefore to adopt tools that are conversant with AFP and other print protocols like Xerox Metacode and LCDS, and that give you the ability to continue to maintain a document archive in the original format while reaching inside that repository to extract important data for analysis, insight and process improvement.

Presentment of information needs to accommodate an increasing variety of formats as well. For example, one common standard is to archive documents in AFP and present them it as PDF. For years PDF has been the common go-to format for viewing online. But as technology has evolved and user habits have changed, PDF has relinquished its reign as the preferable format since an old-school PDF in a mobile environment can be a problem to view. It is important to adopt solutions that are HTML/device aware and give you the ability to render information based upon the device that’s requesting it regardless of how it’s stored within the archive.

Moving Forward

The insight contained in transactional document archives represents a treasure-trove of business intelligence for organizations today. Extracting documents for re-distribution and data from these archives sheds light on important information that has been traditionally unseen and overlooked in corporate decision-making and strategic planning. This dark data, while incredibly valuable, also poses great risk and expense to an organization with regard to court action, legal discovery and regulatory compliance. The key is to have a well thought out strategy in place about what the information means, why it’s important to extract it and use it, and which tools, partnerships and capabilities will be most effective in allowing you to capitalize on the opportunity.

That’s where we come in. At CrawfordTech we are experts in customer communications, transactional print protocols and enterprise content management. While some providers focus on one area, we work with them all and truly take advantage of the opportunities available today. Our award-winning suite of solutions provide the next-generation tools you’ll need to get the job done and with CrawfordTech as a partner, you will be more able to meet the rising demand for increased business intelligence and organizational performance.

Want to learn more? Simply call 1-866-679-0864 for a free executive consultation

CrawfordTech Solutions

Crawford Technologies develops software and solutions to help enterprises optimize and improve the secure and accessible delivery, storage and presentment of their customer communications.

With over 1,800 customers on six continents, CrawfordTech solutions and know-how enable the largest banks, insurers, healthcare providers, utilities and print services companies to use their existing technologies, documents and data in new ways. We help them navigate the challenges in leveraging legacy applications in the platforms and applications of the future.

CrawfordTech’s products, services and domain expertise reside at the nexus of content, data, and output management and are essential components of our customers’ digital transformation, output management and document accessibility strategies.