RIM's Role in

Harnessing the Power of Big Data

Kevin L. Dale, CRM

Organizations can maximize the returns on their big data strategies by giving records and information management (RIM) professionals a seat at the table. Applying RIM principles across the enterprise will help organizations improve the quality of their data and access to it, making big data analysis cheaper, quicker, more efficient, and more accurate.

Organizations are struggling to extract value from their exploding, exponential information growth. Many are turning to big data analytics to mine the gold from the data they store. But are they really positioned to mine for gold using a systematic and controlled approach, or are they simply punching holes in the ground and hoping to strike oil? If they strike gold, are they doing their due diligence to ensure sustainable results?

Too often, RIM is the missing ingredient in this big data conversation.

Technology Is Not Enough

While data analytics software can perform complex algorithms that help predict outcomes from enormous data sets, these tools alone cannot determine if the data being evaluated is of sufficient quality to provide the most accurate results. Organizations often must first make significant investments in preparing the data for analysis, scrubbing or normalizing the data set for missing or bad values.

In fact, Thomas Davenport, a well-

RIM also can play a role as part of the gating process for introducing technology.

known information management and analytics expert, wrote in the *CIO Journal* article "Taming the 'Data Plumbing' Problem," "It is often so difficult to extract, clean, and integrate data that data scientists can spend 90% of their jobs doing those tasks."

The organization has to pay data scientists or analysts a tremendous amount just to determine the structure of columns and rows within poorly defined structured data tables in order to apply the variables to the statistical software. And they may also need to do several costly iterations of mining unstructured data just to identify key repetitive text that can be evaluated for its relevance because the data often lacks sufficient context that comes with good metadata.

Most organizations have a records and information management (RIM) department with skilled professionals who know how to apply RIM principles to the information landscape. They accomplish this by inventorying where all of the organization's data exists, establishing a data map, and understanding the business processes that result in data. Big data can partner with RIM to reduce the costs of identifying, preparing, and analyzing data.

This is why organizations seeking to leverage big data tools to make better business decisions are advised to invite RIM professionals to the conversation.

RIM Reduces Costs, Adds Value

Experienced RIM professionals can help solve big data problems because they know how to organize data to meet RIM requirements, use a naturally consultative approach to help-

ing business units understand how to manage information throughout its lifecycle, and know how to drive improvements – all things that will benefit the analytics team.

By creating synergies between RIM and the big data program staff and integrating RIM principles into all processes that "touch" data, organizations can forge and maintain a sustainable path to the information governance (IG) needed to ensure high-value data. This will reduce costs and add value to the big data program, by:

- Reducing data cleanup and preparation
- Reducing storage costs and risks
- Improving the identification of data through good metadata
- Improving access to data
- Increasing employee efficiency
- Producing better, analytic-driven decision making

RIM Controls 'Four Vs of Big Data'

Through a systematic approach to IG and the application of RIM principles, skilled RIM professionals can help improve the "four Vs of big data":

- Volume – How much data there is
- Variety – The locations and types of data
- Velocity – How much streaming data is being generated
- Veracity – The quality of data

Volume

The more information there is, the more it costs to manage and analyze it. A mature RIM program can help trim data analysis costs by reducing extraneous data through proper retention and disposition processes.

For example, RIM establishes requirements to ensure that appropriate de-duplication occurs in unstructured data repositories. This means the analytics team will have to sift through less poor-quality data, thus lowering the costs of obtaining the data, reducing the time to evaluate it, and enhancing the results since duplicate information won't be included to skew the results.

RIM also can help the big data program be more efficient by ensuring that data in different life cycle stages remains identifiable and retrievable. For example, RIM can identify and locate relevant data in storage locations the analytics team may not know about.

Variety

RIM programs have a consolidated inventory of the repositories where data resides. It also can develop controls that slow down or prevent repository creation, such as collaboration sites and shared drives, thereby mitigating the risk of proliferation of uncontrolled repositories and the associated big data costs.

A systematic approach to allowing new data repositories can help ensure that data is appropriately controlled for information-related risks and that any new repositories are added to the data map so the big data analytics team will always have a current view of everywhere data resides.

RIM also can play a role as part of the gating process for introducing technology. RIM should work with application development and other technology teams to embed RIM requirements for technology that will impact the creation and storage of data.

The procurement team also needs to understand the RIM requirements for new tools that create and store data. It must inform RIM when any tools are decommissioned and ensure that contracts with third-party service or product providers address how the organization can keep control of

its data so it is appropriately retained and managed when a contract expires or a tool is no longer supported

Finally, RIM should enlist the sponsorship of senior leaders to drive the alignment of RIM principles within IT processes. Big data teams should define and embed the role of RIM within their own governance documents to cement this valuable relationship.

Velocity

The proliferation of high-speed data collection sensors is a major contributor to the big data opportunity, and devices that lack sufficient IG are a major contributor to big data risks. Fast data doesn't have to mean *uncontrolled* data, though. Applying RIM principles will:

- Ensure that the data's structure and storage format enhance and maintain the data's value
- Ensure the data is organized and actively managed during all phases of its life cycle
- Reduce the risks posed by the data

RIM teams are invaluable partners for analytics teams because RIM knows how information is flowing through and being used by the organization. This level of business understanding is a key tenet of data science; data scientists need to know if others are adding information to the data, manipulating it, or creating additional repositories.

A RIM and big data partnership can decrease the time it takes data scientists to identify and understand an organization's information flows, which once again drives down big data program costs and improves results.

Veracity

Organizations that are going to invest in a big data strategy should have improving data quality as a primary goal. Poor data quality is an obvious challenge to a big data initiative and a drain on resources. It doesn't make

sense for an organization to hire a big data analytics team, identify poor-quality data repositories, and then do nothing to improve them.

This is where the big data team can help RIM; it can communicate where poor-quality repositories are and recommend practices that will enable the analytics team to harvest better quality data. Examples of poor quality data are:

- That which has little or no metadata
- Duplicate information
- Missing or insufficient data

Applying RIM principles to these repositories will improve the data's content, context, and structure.

reviewing e-mail as evidence is risky because a message can easily be taken out of context, resulting in creative conclusions about its true meaning. The same holds true for data analyzed in the course of a big data project.

Unstructured data by its very nature lacks sufficient context to permit understanding its true meaning without performing a word analysis on common terminology used within it or comparing it with similar documents. Even then, the level of confidence is often tens of percentage points away from 100%.

RIM can establish controls around data that will improve its context. The primary ways are through the

Organizations that are going to invest in a big data strategy should have improving data quality as a primary goal.

The Importance of Integrated Data Governance

Integrating a big data-RIM governance approach with solid RIM principles ensures that new processes, procedures, and technologies are aligned with requirements to manage the data life cycle. This will improve the data's content, context, and structure, allowing the organization to extract big data value by the most cost-efficient means possible.

Improving Data Quality

One way RIM principles can improve the quality of data is to prevent its unauthorized alteration or deletion. The end goal is to have a greater level of confidence in the results of analytics, knowing that they are performed on data sets that come from well-controlled repositories.

Ensuring Data Context

Everyone knows that context matters. During litigation, for example,

labels in structured data tables and the metadata in unstructured data. RIM must be empowered to drive standards for the labels and metadata of all organizational data that can help data analytics teams. For example, it should include fields that identify data owners so the big data team can quickly get from the owners any needed permissions to access their data.

Defining Data Structure

By working with the big data team, RIM can establish a governance approach to drive the design of file plans that will structure data and metadata so the data is easily accessible to be pulled into analytics software, facilitating its analysis. Empowering RIM professionals to establish enforceable standards for file plans will provide a repository structure that reduces risk, decreases storage costs, and improves analytic alignment.

Implement an enterprise-wide policy that defines the roles and responsibilities for each employee.

File plans for unstructured data are usually developed using records management and enterprise content management systems, although data may also reside within shared drives and collaboration environments that lack any coherent file plan. This means the storage and file plan structures for data should be aligned with content management and data analysis software requirements where possible.

How to Align Big Data and RIM Governance

One way to begin a big data-RIM alignment is to establish two groups – an IG council that makes strategic decisions on standards for driving big data-RIM requirements throughout the organization and prioritizes work, and a big data-RIM workgroup that is responsible for doing the work.

Establishing an IG Council

The IG council should develop and advance the strategic goals of the big data and RIM program throughout the organization. It should be composed of senior leaders from RIM, IT, information systems, legal, enterprise risk, audit, compliance, governance, and finance. It might also include members from sales, marketing, customer service, and communications if, for example, the nature of big data strategy involves predictive analytics to drive sales.

The IG council must be empowered to decide how data is governed to mitigate RIM-related risks in parallel with enhancing the effectiveness of data analytics. For example, the council would ensure that as poorly structured data sources are identified, the resources to mitigate them are quickly made available.

The council must be visible to the entire organization and consistently communicate the importance of managing data as an asset.

Establishing a Big Data-RIM Workgroup

A big data-RIM IG workgroup should be made up of front-line management from RIM, data analysts, and key IT and business partners. It is responsible for the tactical functions the IG council has mandated and prioritized.

This group aligns managing information with the needs of the analytics team and the business units, is tasked with accomplishing initiatives in a timely manner, and reports its progress to the IG council on a regular basis.

Institutionalizing the Approach

Establish charters for the IG council and the big data-RIM governance workgroup, clearly identifying each group's mission, vision, purpose, roles and responsibilities, goals, and metrics. Institutionalize their activities through documented governance.

Implement an enterprise-wide policy that defines the roles and responsibilities for each employee. Reaffirm the policy each year and provide refresher training.

Each business unit should contribute to the transparency of the RIM programs by establishing and enforcing a procedure for communicating new repositories to RIM. They should also perform quality assurance on and oversight of the data they produce, use internal metrics in validat-ing their quality control programs, and assign liaisons to ensure that business-level governance is followed.

Benefits of a Big Data-RIM Governance Framework

Adopting an integrated big data-RIM approach gives organizations a competitive advantage over organizations using big data analytics alone because the latter approach does not lend itself to the types of sustainable improvements necessary to continue to enhance the quality of and access to data.

A big data-only approach will also fail to reduce information-related risks. In time, the costs and risks associated with poor-quality data residing in uncontrolled repositories will eat away at any competitive advantage and become a significant liability. The costs of delaying a big data strategy that includes RIM principles will also increase exponentially with data growth.

Organizational leaders should aggressively seek to mitigate these risks and rethink their big data strategies to include RIM principles. By making this shift, they will reduce poor-quality data and increase their confidence level in their big data analysis results.

Better results should lead to better decision making and increased profits. Employees and the data analytics team will benefit from their ability to identify and access data in a timely manner, which will improve organizational efficiency in the use of its data while driving down information-related risks.

By leveraging a big data-RIM approach to establishing requirements to control the proliferation of information, the organization will reduce the costs associated with storing, mining, and analyzing that data. And that's just good business. **END**

Kevin L. Dale, CRM, can be contacted at kevin.dale@chi.frb.org. *See his bio on page 47.*