

**Whitepaper**

# QUANTIFIED INFORMATION GOVERNANCE:

A NEW PATH TO VALUE FROM DATA

Barclay T. Blair, Executive Director,  
Information Governance Initiative

**GUIDANCE**  
SOFTWARE 

 information  
governance  
initiative

## **ABOUT THIS SPECIAL REPORT**

An Information Governance Initiative (IGI) Special Report is an in-depth analysis of an information governance (IG) topic of special interest to the IG community that is created with the support and collaboration of an IGI Supporter. This Special Report was created in collaboration with – and published by – Guidance Software as part of the IGI's mandate to advance the practice and profession of IG. As part of the IGI's commitment to excellence and to maintain objectivity, it does not recommend, evaluate, or endorse specific products, services, or providers.

This Special Report was made possible by Guidance Software's support of the IGI. IGI Supporters like Guidance Software provide an annual financial contribution to support the activities of the IGI, and by doing so demonstrate their commitment to the advancement of IG. The IGI is a think-tank, community and consortium dedicated to advancing the practice and profession of information governance. Join our community today at <http://iginitiative.com>

## **MESSAGE FROM GUIDANCE SOFTWARE**

This Special Report explores the idea of “quantified information governance” - a concept by IGI that addresses both the philosophy and discipline of intelligent software to enable data-driven or evidence-based IG decision making. Guidance Software's new product - EnForce™ Risk Manager – is designed to systematically identify, categorize, and remediate sensitive and aged data throughout the enterprise. This product directly applies to the concept of quantified information governance as defined by IGI. Please visit <https://guidancesoftware.com/enforce-risk-manager> to learn more about EnForce™ Risk Manager.

## OVERVIEW

Our research at the Information Governance Initiative overwhelmingly shows that information governance (IG) is shaping the management practices of organizations across industry sectors as a new approach to generating value from data while simultaneously addressing risk. In fact, most organizations doing IG have multiple projects in flight, and they are spending real money to get them done (on average, large organizations have seven active IG projects each costing over \$750,000 USD).<sup>1</sup>

However, our research also shows that IG projects are difficult to get off the ground, ill defined, and simply take too long to finish (35% of IG practitioners say it takes “longer than a year” to get a project started after the decision to do it has been made).<sup>2</sup> The reason for this is clear: most organizations have yet to structure their organization in a way that supports IG project funding and execution. But, even this is changing as organizations are appointing new leaders like the Chief Information Governance Officer (CIGO) who play a new coordinating role for the multitude of information-related challenges that organizations face – from privacy and cybersecurity to monetization and compliance (67% believe a CIGO is essential to IG success).<sup>3</sup>

At the same time, we are experiencing a generational shift toward the use of data to drive decision making at both the most trivial level of our personal lives and at the most consequential levels of global financial policy and multibillion-dollar business bets. It has been hard to miss the hype around big data, but amidst the hype, real transformation is happening. For example, despite some who cynically dismiss the “Internet of Things” (IoT) as toasters and blenders uselessly connected to the Internet, GE recently reported over \$1B USD in revenue from its IoT software business – hitting that mark faster than any GE division in history, and only a small fraction of what their CEO is betting it will become.<sup>4</sup> Even regulators like the Federal Trade Commission have weighed in on the regulatory implications of IoT and are grappling with it from a policy perspective, advising companies to only retain “information for a set period of time, and not indefinitely.”<sup>5</sup>

## QUANTIFIED BY DATA: LIFE, BUSINESS, AND GOVERNMENT

Life, business, and government are increasingly quantified by data – data that is driving critical decision-making. The demand for devices that track and analyze the data we generate just by living demonstrates the rise of the “quantified self.”<sup>6</sup> Analysts predict that by 2020, the market for “fitness wearables” will grow to \$10B USD (from \$3.3B USD),<sup>7</sup> with over 100M people using the devices to enable data-driven decisions about health, sleep, and exercise.<sup>8</sup>

The promise of data-driven decision making is this: processing and analyzing data at a scale far exceeding the capabilities of the human brain will transform our ability to understand and predict reality. The IGI believes that ability to govern information in a way that enables these deeper insights, unforeseen efficiencies, and new business models is what will separate the winners from the losers in this new era.

But we still have a long way to go. While we invest in technology that can beat a human at Jeopardy® in one part of our organization, we are stuck with the technology that prints Trebek’s cue cards in another. For all the big data sexiness, “up to 80% of the total development cost of an analytics project” is spent on “data discovery and wrangling . . . the most tedious and time-consuming aspects of an analysis.”<sup>9</sup>

Why does it take so long? Because most organizations quite simply have very little idea what data they have, where that data lives, what the data means, what rules must attach to the data, and whether or not the data represents measurable value or risk. Consequently, our data is messy, incomplete, difficult to find and access, duplicative, and missing context essential to enable its analysis and use. In short, it is the inevitable outcome of a generation of attempting to force analog practices to work in a digital world. It has failed.

Most organizations continue to make management decisions about their information based on *tradition*, *superstition*, and *supposition* instead of *innovation*, *evidence*, and *analysis*.

It’s time that our approach to governing our information caught up to the information age. It’s time for a new idea. We believe that this new idea is **Quantified Information Governance**.

## WHAT IS QUANTIFIED INFORMATION GOVERNANCE?



*“Former Bank of England Governor Mervyn King last month blamed the Riksbank’s poor forecasting record on policy makers putting too much faith in computer-generated models and not enough in their own gut.”<sup>10</sup>*

*Bloomberg Business, February 2016*

Quantified IG is the application of smart technology and evidence-based practices to the governance of information. It ensures that we have essential **facts** about our information and our operating environment so we can make **evidence-based decisions**.

Quantified IG requires technology that supports the gathering and analysis of facts about our information. Just as the big data revolution is in part possible because of the availability of newly-affordable Internet-scale infrastructure, Quantified IG is made possible by a new generation of file analysis, indexing, and analytics technologies (many honed on the whetstone of high stakes/high data volume/high scrutiny lawsuits and investigations) that promise to profoundly increase our insight into unstructured information (i.e., generally, any information that does not reside in a database).

The idea that we should make decisions based on facts or evidence of course derives from the Enlightenment and the scientific method itself. But even in areas where you might expect that this approach is already baked in, it is still challenging. In the 1990s, for example, the concept of “evidence-based medicine” was introduced into the medical field and has since taken hold as an operating philosophy in branches of medicine from optometry to dentistry. Evidence-based medicine is defined as: *“the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients.”<sup>11</sup>*

If the practice of medicine – which has embraced the scientific method for over a century – can benefit from a heightened focus on evidence-based decisions and policy, then surely there are other practices that could benefit from it as well. Any come to mind?

Why do we have 1000 categories in our records retention schedule? Because that’s the way the last guy did it. Or, because we inherited the schedule from a company we acquired. Because Janice liked it that way. Because that’s what makes the most sense to me. Because that’s what my old boss told us to do. Because that is what the consulting company sold us.

But is it right? Is it true? Is it the best way? Are these justifications based on anything more than tradition, superstition, or office politics?

Without quantification, it is impossible to know.

## QUANTIFIED IG APPLIED: DATA REMEDIATION

### Remediation is a Business Decision

Many practitioners mistakenly believe that data remediation decisions (e.g., keeping, deleting, or migrating information) are legal or technology decisions. But, they are actually business decisions. What is the difference and why does it matter?

The difference is that legal analysis and technology evaluation are only part of what is required to make a remediation decision. It matters because the way we perceive remediation clearly correlates with the quality of our approach and our level of success.

A business decision is the act of:

- Gathering the best available facts
- Calculating benefits and costs
- Assessing risk



Armed with this insight and knowledge about our organization's capabilities, goals, and values, an outcome is projected, a decision made, and a course of action set. Of course in practice, only the most consequential business decisions are made with this level of formality, and time, budget, politics, and other factors mean that the decisions are necessarily imperfect.

However, this framework clearly illustrates why we have struggled with remediation: we have lacked high quality, practical, and cost-effective tools and techniques to make good business decisions about our data. Quantified IG will change that.

### Quantify Risk to Act on Insight, Not Fear



*“So much of what we think and do about risk does not make sense. In a 1990 paper researchers . . . compared people’s levels of concern about risks with objective measures of those risks. The results can only be described as scrambled. In some cases, concern rose and fell as the risk rose and fell. In others there was ‘wild fluctuation’ in the level of concern that had absolutely no connection to the real risk. ‘There is no generally applicable dynamic relationship between perceived and actual risk,’ the researchers politely concluded.”*

**Dan Gardner, Risk: Why We Fear the Things We Shouldn’t, and Put Ourselves in Greater Danger<sup>12</sup>**

In 2002, Daniel Kahneman (a psychiatrist, not an economist) was awarded a Nobel Prize in Economics. Why? For his groundbreaking work on the psychology of judgment and decision making, proving that human beings are riddled with cognitive biases and flaws, one of which is that we are quantifiably bad at quantifying risk. We consistently overestimate some risks and underestimate others. We wildly overestimate the risk of dying in a plane crash and wildly underestimate the risk of dying in a car crash on the way to the airport when in fact the data is very clear that our worries should be focused on precisely the inverse.

Our research tells us that three things bedevil most data remediation projects.

#### **1. Lack of insight into the information environment.**

We discussed above how a new generation of software is solving that problem, and we strongly advocate that IG practitioners investigate and evaluate those technologies.

#### **2. A breakdown in corporate governance**

that fails to make clear who actually has the mandate and authority to make remediation decisions.

#### **3. Fear.**

Decision-makers of all stripes are guilty of making fear-based decisions about information. For example, they are afraid that, if they throw the wrong thing away, they will face sanctions, fines, or even jail. So, they decide to do nothing.

This fear is the predictable result of a decade of education and awareness centered on the “CEO in prison stripes” image derived from Sarbanes-Oxley’s obstruction of justice provisions or the “you accidentally recycled a backup tape, so you lose” sanctions feared in civil litigation after the 2006 revisions to the US Federal Rules of Civil Procedure.

Is this fear evidence-based? Are organizations making data-driven decisions about the fate of their information, driven by real insight into their information environment and quantification of the risk and reward?

Our research suggests the answer is no. In fact, most organizations’ data remediation plans are stalled or moribund because of a failure to bring the evidence-based practices they are using in other parts of their business to bear on the problem of IG generally, and data remediation specifically.



## Quantifying Remediation Risk



*"I am not aware of any study that indicates that such [spoliation] motions are relatively common."*

**Emery G. Lee III, US Federal Judicial Center**

Many executives seem to imagine that the likely (and perhaps inevitable) consequence of getting rid of useless data is being hauled up on the stand in a courtroom to defend his or her terrible act of "spoliation" (i.e., the illegal destruction of evidence).

But how likely is this?

Of course, the facts and context of any lawsuit are ultimately determinative, but the historical data is very instructive.

First of all, studies indicate that in the USA, **less than 2%** of all federal civil cases even go to trial,<sup>14</sup> and within that tiny percentage of cases, only a small portion see motions even alleging spoliation (0.15% according to one influential study,<sup>15</sup> which also found motions related to electronic evidence were actually granted only 23% of the time).<sup>16</sup> And in that tiny slice of slices the number of cases resulting in spoliation sanctions is barely measurable. In fact, such cases are so unusual that they are studied by lawyers – like biologists studying a new species – and widely referred to even by non-lawyers by name.

Similarly, a recent 30-year study of the US federal civil justice system found that there were only 230 cases in which any kind of sanction related to e-discovery was imposed. In only 136 of these cases was spoliation actually alleged.<sup>17</sup> Although the overall number of cases involving e-discovery has obviously been accelerating in the past 30 years, the data shows that the frequency of spoliation awards for digital information is still very low.

There are certainly caveats to this data. It does not account for e-discovery battles occurring pre-trial or outside of the courtroom. It does not reflect the fact that, for certain kinds of cases and certain fact patterns, the probability of a successful spoliation claim (or just a spoliation claim that makes a case longer and more expensive) approximating the executive's nightmare vision skyrockets. And there are other limitations.

However, the data is clear: decision makers approving data remediation as a routine part their IG program are extremely unlikely to find their courtroom fears realized. Despite this truth, there remains an unhealthy, inertia-causing gap between the perceived risk of bad things happening because of data remediation and the reality. The good news, though, is that every dollar smartly invested in IG can drive the quantifiable risk down and even move it towards zero.

## It's Getting Better



*"The changes to FRCP Rule 37(e) now focus the court's attention on the most important aspect of spoliation, which is prejudice. The new rule rejects the negligence standard, which means that organizations no longer need to fear case-ending sanctions for the almost inevitable loss of information that happens simply as an ordinary part of operating a business. This empowers us to move forward with data remediation projects without fear that their good faith efforts to govern information will result in severe sanctions."*

**Bennett B. Borden, Chief Data Scientist,  
and Chair of the Information Governance and eDiscovery Group at Drinker Biddle & Reath**

Quantified IG empowers organizations to make high-quality business decisions about their data, even in difficult business environments. Thankfully, in at least one aspect of that environment, things are getting better.

Numerous amendments to the US Federal Rules of Civil Procedure (FRCP) went into effect on December 1, 2015. Rule 37, and in particular Rule 37(e), defines when and how courts can address digital information and data that was not preserved. Under the prior Rule, the severity of a potential spoliation sanction relied heavily upon the level of culpability of the alleged spoliator. The cases interpreting the Rule had adopted a negligence standard and awarded sanctions based upon the level of culpability. However, under the new Rule 37(e), this negligence standard has been abandoned. Instead, the courts now look to the degree to which the non-spoliating party was prejudiced by the loss of information.

Critically, only upon a finding of prejudice may a court award sanctions, and then “may order measures no greater than necessary to cure the prejudice.” The state of mind of the alleged spoliator is irrelevant to this analysis except in those cases where the court finds “intent to deprive another party of the information’s use in the litigation.” Only upon this finding may the court apply the severe sanctions such as an adverse inference or dismissal of claims or the entire case. Thus, absent extreme cases of willful spoliation, the most severe sanctions are no longer in play.

This is great news for organizations adopting Quantified IG.

### **TAKING ACTION: GETTING STARTED WITH QUANTIFIED IG**

You need data to get started with Quantified IG. You may be thinking, “Huh? The last thing I need is more data.” You probably do have too much data. But, how much data do you have about your data?

If you are like most organizations, you are starved of data that truly provides insight into your information environment. What do you know about it for sure? You might have a general sense of how much storage is allocated to messaging, your file shares, and a few key business applications, but what else? Where is your most valuable data and what is it? What is your most valuable data? How many different places do private and sensitive data live? Which information should you spend more time and money governing, i.e., identifying, organization, and protecting?

Can you answer these kinds of questions in a way that is clear and convincing enough to drive multimillion-dollar business decisions? Or drive decisions that have major career-limiting implications if wrong? If you have insight into your data, do you actually believe in that insight deeply enough to take action? Our research shows that most organizations do not.

This is the role of Quantified IG. So, how should you get started?

## How Quantified IG Powers Data Remediation: Three Steps to Actionable Insight

### Step One: Facts



Quantified IG for remediation projects relies upon accurate and believable insight into your information environment. This means discovering the facts about your data that everyone in the decision chain will believe in strongly enough to make consequential decisions. Today, we have new and practical ways to do this. Just as we all know that the human generated food log is more likely to reflect the moral fiber of its creator than the actual fiber its creator consumed, a data inventory or dashboard created using a process and technology that is objectively high-quality will generate insight that is a more accurate reflection of reality, and thus actionable.

As such, you need tools and techniques for poking and prodding into all the places where your information lives. In most organizations, there is no central inventory of those locations, or even a map to find them, so self-discovery is a critical capability. In addition, the ability to maneuver within the security protocols protecting those repositories is critical if you want a complete picture.

### Step Two: Analysis



Quantified IG requires tools and techniques that enable not only the discovery – but also analysis – of facts. Actionable insight depends upon both the quality of the analysis and on the presentation of the analysis. The goal is simplicity and clarity with a bias towards the language of business (as opposed to technical or statistical language). The bigger the decision, the simpler and clearer the presentation of the analysis must be. Your analysis must be supported by a deep and defensible set of practices and decision-making criteria. For example, if you are claiming that 30% of the information in your departmental shared drives is duplicative, then be prepared to explain the methodology you used to determine what actually counts as a duplicate and why that is the right methodology for your organization. Are two documents that are identical in every way except their file name duplicates? What if they are identical in every way other than their location or the name of their creator? Can duplicates be remediated without breaking a business process or infuriating a user? These are questions that need to be confronted and answered, even if they don't make it into the dashboard view.

### Step Three: Action



Quantified IG has a bias towards action. The best facts and analysis mean nothing (and in fact may create new risks) without action. You need an articulated, funded, and managed plan. You need the technological capability to take action based on the evidence that you have collected and created. The value of a “dashboard” is only realized when the business decisions that the dashboard enables can actually be executed. This can be very challenging in the unstructured information environment, which typically has very little of the governance that would drive consistency in management retention, protection, and classification practices, so you must look for tools with the capability to operate in this chaotic environment.

## **QUANTIFIED IG: A NECESSARY EVOLUTION FOR IG**

There is a generational movement to use quantification (evidence analyzed for insight) to inform a growing spectrum of decisions in our world. At its heart, this movement continues the intellectual evolution that began in the Enlightenment. The promise is better decisions: decisions based on better information and evidence; decisions that are more likely to be correct; decisions that are more likely to result in the planned outcome.

However, most organizations are only beginning to adopt this approach for the governance of their information. Many decisions in IG are made based on nothing more than a cognitively suspect human calculation of risk. Or, very commonly, decisions are effectively not made, evidencing the bias towards inaction that plagues organizational decision making regarding information. There is no excuse for this to continue. The techniques of data-driven decision making and quantification are well understood. We now have exciting new technologies that finally empower us to collect facts, conduct deep analysis, and at last, to take action on our data. In a world where most organizations are experiencing exponential growth of data generally, and pools of dark and potentially dangerous data specifically, a commitment to action, driven by powerful emerging IG best practices and technologies, is the only way forward.

## CITATIONS

- <sup>1</sup> Information Governance Initiative, "Information Governance Initiative Annual Report 2015-2016" (Information Governance Initiative LLC, September 2015).
- <sup>2</sup> *Ibid*, 23.
- <sup>3</sup> *Ibid*, 39.
- <sup>4</sup> Quentin Hardy, "G.E. Opens Its Big Data Platform," *New York Times*, October 9, 2014.
- <sup>5</sup> Federal Trade Commission, "FTC Report on Internet of Things Urges Companies to Adopt Best Practices to Address Consumer Privacy and Security Risk," January 27, 2015. (Refers to: Federal Trade Commission, "Internet of Things: Privacy and Security in a Connected World," FTC Staff Report).
- <sup>6</sup> "Quantified Self (QS)" was developed into a "movement" largely by two editors from *Wired Magazine*. (Deborah Lupton, "The Quantified Self Movement: Some Sociological Perspectives," (November 4, 2012). <https://simplysociology.wordpress.com/2012/11/04/the-quantitative-self-movement-some-sociological-perspectives/>
- <sup>7</sup> James Moar, *Wearables-The Heartbeat on Your Sleeve* (Juniper Research, November 2015), 4.
- <sup>8</sup> James Moar, *Fitness Wearables-Time to Step Up* (Juniper Research, January 2016), 4.
- <sup>9</sup> Victoria Louise Lemieux, Brianna Gormly, Lyse Rowledge, "Meeting Big Data Challenges with Visual Analytics," *Records Management Journal*, July 2014.
- <sup>10</sup> Johan Carlsrom, "One Theory on Sweden's Inflation Forecasting Failures: It could be a deep-rooted bias that's holding the Riksbank back, SEB says," *Bloomberg*, February 15, 2016.
- <sup>11</sup> DL Sackett, WMC Rosenberg, JAM Gray, RB Haynes, WS Richardson, "Evidence Based Medicine: What It Is and What It Isn't," *BMJ* 312 (1996): 71-2.
- <sup>12</sup> Dan Gardner, "Risk: Why We Fear the Things We Shouldn't, and Put Ourselves in Greater Danger," *Emblem Editions* (January 27, 2009).
- <sup>13</sup> 18 U.S. Code § 1519 - Destruction, alteration, or falsification of records in Federal investigations and bankruptcy.
- <sup>14</sup> "1.8% of federal civil cases went to trial. Patricia Lee Refo, "The Vanishing Trial," *Journal of the Section of Litigation, ABA* (Winter 2004), p.2.
- <sup>15</sup> Emery G. Lee III, *Motions for Sanctions Based Upon Spoliation of Evidence in Civil Cases* (Federal Judicial Center, 2011), 4.
- <sup>16</sup> Emery G. Lee III, *Motions for Sanctions Based Upon Spoliation of Evidence in Civil Cases* (Federal Judicial Center, 2011), 8.
- <sup>17</sup> Dan H. Willoughby, Jr., et al., "Sanctions for E-Discovery Violations: By the Numbers," *60 Duke L. Rev.* (2010).



### **ABOUT GUIDANCE**

Guidance exists to turn chaos and the unknown into order and the known-so that companies and their customers can go about their daily lives as usual without worry or disruption, knowing their most valuable information is safe and secure. The makers of EnCase®, the gold standard in forensic security, Guidance provides a mission-critical foundation of market-leading applications that offer deep 360-degree visibility across all endpoints, devices and networks, allowing proactive identification and remediation of threats. From retail to financial institutions, our field-tested and court-proven solutions are deployed on an estimated 33 million endpoints at more than 70 of the Fortune 100 and hundreds of agencies worldwide, from beginning to endpoint.

Guidance Software®, EnCase®, EnForce™ and Tableau™ are trademarks owned by Guidance Software and may not be used without prior written permission. All other trademarks and copyrights are the property of their respective owners.