Data Intelligence in Context: Enabling Data Governance for Digital Transformation

December 2018

Written by: Stewart Bond, Research Director, Data Integration and Integrity Software

Introduction

Maximizing success in digital transformation (DX) requires organizations to be enabled by data, not restricted by data governance. But as the need to leverage data escalates, so too does the need to protect data within complex ecosystems as well as business, IT, and data environments. Data governance is no longer optional for enterprise organizations.

Aside from complying with new regulations, organizations are finally realizing the value of data as an asset that needs to be protected, managed, and maintained to increase asset value.

Data governance efforts are still being affected by a lack of intelligence about data. Organizations that lack data knowledge gained through data intelligence are wasting up to 30% of their time on data governance activities. Data intelligence software can help through data discovery, cataloging, profiling, mastering and providing lineage across the data supply chain. Chief data officers, chief information officers, IT and line-of-business leaders, data security officers, data stewards, and data owners who are informed by data intelligence will gain the knowledge required for their organization to be enabled by data.

Business and IT have always been at a divide when it comes to data ownership, and data governance is too often associated with "no," not "yes." The "no" of data governance can be turned into the "yes" of data enablement. For organizations to be enabled by data, the right data needs to be trustworthy, timely, and delivered to the right resources at the best time. Data enablement still requires governance to infuse trust, timeliness, and availability and provide protection, but not at a level that stifles innovation and exploration, or innovation will focus on how to get around data governance.

Many organizations with failed data governance attempts can point to outdated or absent data intelligence that is captured and maintained manually. A recent IDC survey discovered that spreadsheets, custom software, documents, and word of mouth are among the most frequently used methods of cataloging data. Automation of data discovery using machine learning, where the system is trained by people and metadata, is the only possible method of gathering data intelligence in this new era of DX. Data intelligence software is required to inform the discipline of data governance,
represented by a collection of automated capabilities and crowdsourced knowledge that helps organizations answer the following questions:

» **Who** can access the data, and who is using it?

» **What** does the data mean?

» **Where** is the data, and where did it come from?

» **When** was the data last validated?

» **Why** does the data exist?

» **How** can the data be consumed?

Data intelligence software also helps answer a new question: What are the relationships inherent in data?

**Why Data Intelligence for Data Governance Is Important**

Data alone does not distinguish the enterprise — how the enterprise is enabled by data is the differentiator. Data enablement requires governance and governance requires data intelligence to deliver the right data to the right resource at the best time.

The first and obvious measure of data governance ROI is avoidance of regulatory fines. There has been a lot of focus on the General Data Protection Regulation (GDPR) because if an organization is found to be noncompliant with the GDPR, it can be fined up to 4% of its annual revenue. The GDPR not only introduces new strict compliance requirements on data but also is the first regulation to threaten such large fines for noncompliance. Regulations have emerged for multiple reasons, but first and foremost, they are for protection of data and sensitivities that lie with the data. Enterprises complain about having to comply with regulations because they cost time and money, but if the "bad guys" have figured out how valuable information is, enterprises also need to value their data and invest in capabilities to protect it. The by-products of complying with emerging regulations will be not just fine avoidance but also more efficient and effective business operations and employee productivity because an increase in data intelligence can increase data literacy and knowledge.

Figure 1 illustrates the results of a recent IDC survey, demonstrating that the 80/20 rule is still prevalent across organizations; that is, 80% of time is spent on data discovery, preparation, and protection, and only 20% of time is spent on actual analytics and getting to insight. Data intelligence software has the potential to change this ratio, giving users the ability to find data easier and understand context and definitions for better integration, resulting in more time for analysis.
FIGURE 1: *Inefficient Data Activities*

Q. How many hours per week on average do you spend on each of the following data-related activities?

% of Time Spent on Data Activities (Weekly)

- Searching: 20%
- Preparing: 37%
- Protecting: 24%
- Managing data: 81%
- Analyzing: 19%

\[ n = 300 \]

Source: IDC's Data Integration and Integrity End-User Survey, November 2017

The survey also asked people how often they are successful at completing data-related activities. Their answers gave us the inverse; that is, how often they are unsuccessful: an average of 12 hours a week (see Figure 2). People are also wasting an additional 10 hours per week creating new analytic assets only to learn that a similar asset has already been created by someone else.
FIGURE 2: Time Wasted per Week

Q. How often are you successful in data asset search, preparation, and protection activities?

<table>
<thead>
<tr>
<th>Activity</th>
<th>Time Wasted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preparing</td>
<td>5.7</td>
</tr>
<tr>
<td>Protecting</td>
<td>3.8</td>
</tr>
<tr>
<td>Searching</td>
<td>2.7</td>
</tr>
</tbody>
</table>

Average hours per week

n = 300

Source: IDC’s Data Integration and Integrity End-User Survey, November 2017

In this new era of DX, the number of different types of data is growing, as is distribution of data across hybrid and multicloud environments. However, this data growth is compromising data controls, and a lack of data literacy among users creates a kind of chaos that impacts productivity. Increasing regulatory pressures and labor-intensive processes that cost both time and money are additional challenges that hamper an organization’s ability to govern data and information.

Automating collection of data intelligence and putting it into the hands of professionals who work with data — an increasing number of different roles — can increase data literacy and improve data knowledge. When data enablement through governance is implemented properly, demonstrating regulatory compliance becomes easier, the integrity of data improves, and the organization becomes more effective and efficient. An increase in data intelligence will provide professionals who work with data more opportunities for collaboration and socialization of data, further increasing the knowledge required to be more productive in data-related activities. Improving the efficiency and effectiveness of these professionals, even by a small margin, could have a significant impact on productivity, analytics, and business outcomes.
Data Intelligence in Context: Enabling Organizations with Data

Organizations that approach data enablement and governance with a technology-only solution, thinking it is a technology-only problem, will fail. Data governance is an organizational problem that requires an organizational solution, aligned with corporate strategies and implemented through people, processes, policies, architecture, and technology. Data intelligence software is technology that supports data enablement through governance and is delivered in metadata management, data lineage, data catalog, data profiling, mastering, and stewardship software. Analytics of the metadata managed by data intelligence software will provide new opportunities for insight into the value of data as well as optimization of data enablement and business operations.

Data cataloging software provides the ability to discover and harvest data using machine learning, artificial intelligence, network crawlers, and metadata ingestion. Once data is discovered, catalogs maintain data location, usage information, business glossaries, cross-domain relationships, and crowdsourced collaboration. Data lineage software is focused on tracing the lineage of data used in data integration and business intelligence (BI) reporting, including the ability to map flows in the data supply chain. Data profiling and stewardship software captures schema, content heuristics, statistics, constraints, and data quality parameters such as amount of duplication, levels of completeness, timeliness, and consistency. Data mastering software assists with the definition and structure of master and reference data schema and implements match reconciliation and survivorship rules while keeping location reference information of master data distributed across an enterprise. Functionality delivered by these software solutions supports use cases such as data enablement through governance, data quality management, and data self-service.

Data intelligence is also uncovering data relationships; relationships in the data itself and relationships with enterprise architectures and processes as well as the people who execute them. This is putting data intelligence in context, adding to the knowledge that is required for data enablement.

Data intelligence also can improve data operations; it will inform decisions on access control, authorization and usage policies, encryption, masking, and protection; where to locate data; and how and where replication needs to occur on high-value data assets to maintain business continuity. Data intelligence can also be used to improve the state of data quality because it assists data stewards with understanding data dimensions and shapes and uncovers data duplication, inconsistency, incorrectness, and attributes of timeliness and completeness.

Considering erwin Inc. for Data Intelligence in Context

erwin Data Modeler (DM) has been a choice for chief information officers, chief data officers, data architects, and other IT professionals for more than 30 years. erwin was divested from CA Technologies in 2016, purchased by Parallax Capital Partners with the intent to build a company focused on helping organizations be successful with data. Subsequently, erwin acquired Corso, an enterprise architecture modeling company; Casewise, a business process modeling company; and more recently, AnalytiX DS, a metadata management and data governance company. The integration of these companies and technologies has resulted in the erwin EDGE platform, a data intelligence solution centered on data governance in the context of enterprise architecture, data models, and business processes.
The erwin EDGE platform is a persona-based solution that can be used by IT specialists, data integration developers, data scientists, data stewards, data owners, line-of-business analysts, compliance and security officers, and data and corporate executives. Each persona is given a different interface depending on the types of activities he or she performs with data. The individualized interface provides data intelligence in context to inform the user, giving each the data literacy and knowledge to be more successful working with data. The EDGE platform also provides a collaboration platform to keep the lines of communication open and share data knowledge across the population of data producers, consumers, and stewards.

With this uniqueness of integrating data engineering and governance with business process and enterprise architecture models, erwin has also developed a seven-step method to better outcomes:

1. **Discover data** by identifying and integrating metadata across data silos.
2. **Harvest data** by automating the collection of metadata and consolidating it in a single repository.
3. **Structure and deploy data sources** by connecting physical metadata to business terms and definitions and design standards.
4. **Analyze metadata** to understand how data relates to the business.
5. **Map data flows** to identify where and how data flows across the enterprise and transformations that occur.
6. **Govern data** using a model to manage standards and policies and set best practices.
7. **Socialize data** by enabling stakeholders access to data in the context of their roles.

The erwin EDGE platform is enabling organizations with data, not only in the data intelligence that is being delivered but also in its ability to generate data integration and management code that can be leveraged in data migration and operations. This is helping organizations leverage data from legacy platforms into distributed hybrid and multicloud environments while also being able to understand the impact of changes across the data supply chain as it relates to how the enterprise conducts business.

Organizations that select and implement erwin EDGE as the platform on which to build their data intelligence in context will have an opportunity to change the 80/20 rule by providing users with the knowledge needed to be more effective and efficient with data.

**Challenges**

erwin is forging its way into a market that is getting a lot of attention and expanding, but it’s also a market with some key leaders in mindshare and intellectual property that can be leveraged for deep capabilities in data intelligence. This is a challenge for erwin. As part of CA Technologies, erwin was hidden among the company’s many assets until divestiture in 2016. Thus erwin isn’t at the top of the list for mindshare as a modern data governance solution. Acquisitions made by erwin are what make it unique among all the other data intelligence and governance software players, in that the integration of data, enterprise architecture, process, business glossary, and data dictionary models is a key differentiator. To get ahead of the competition, erwin will need to lean on this differentiator and demonstrate integration of acquired assets to deliver functional and efficient solutions.
Conclusion

Not moving forward with data enablement, made possible by application of data intelligence technology in support of data governance, is a risky venture. Getting the wrong data to the wrong person at a bad time would be the opposite of data enablement, and it could be very problematic. Beyond the risk of incurring significant noncompliance penalties for organizations, data breaches can have a significant impact on company security and public reputation. Low-quality data used out of context could result in bad business decisions, and DX innovation will be squelched, resulting in status quo business operations that may cause declining growth and ultimately failure.

IDC believes that the data intelligence software market will continue to grow, and to the extent that erwin can address the challenges described in this paper, the company has a significant opportunity for success.
erwin EDGE
The erwin EDGE platform delivers an "enterprise data governance experience," so the modern business can accelerate the transformation of mission-critical data into accurate and actionable insights. It ensures collaboration between IT and the business to discover, understand and unlock the value of data both at rest and in motion. With data governance as the driving principle and underlying technology, organizations can gather data intelligence, conduct IT audits, and ensure regulatory compliance. In addition to helping mitigate data-related risks, the erwin EDGE fuels an automated, high-quality and real-time data pipeline that can be used to accomplish any organizational objective — from creating new products and services to increasing customer satisfaction, both of which increase topline revenue. With a role-based approach, data assets can be understood in a business context, their physical existence and lineage tracked, and their security, quality and value maximized.

Website: www.erwin.com

Video Demo: https://www.youtube.com/watch?v=v3q_4PzwppE&feature=youtu.be

About the analyst:

Stewart Bond, Research Director, Data Integration and Integrity Software

Stewart Bond is Research Director of IDC’s Data Integration and Integrity Software service. Mr. Bond’s core research coverage includes watching emerging trends that are shaping and changing data movement, ingestion, transformation, mastering, cleansing and consumption in the era of digital transformation.