CUTTER Business Technology Journal

Management, Innovation, Transformation

Vol. 31, No. 9, 2018

Building a Digital Business STARTS WITH DATA

Barry Devlin
Guest Editor

CUTTER Business Technology Journal

As business models for creating value continue to shift, new business strategies are constantly emerging and digital innovation has become an ongoing imperative. The monthly *Cutter Business Technology Journal* delivers a comprehensive treatment of these strategies to help your organization address and capitalize on the opportunities of this digital age.

Cutter Business Technology Journal is unlike academic journals. Each monthly issue, led by an expert Guest Editor, includes five to seven substantial articles, case studies, research findings, and/or experience-based opinion pieces that provide innovative ideas and solutions to the challenges business technology professionals face right now — and prepares them for those they might face tomorrow. *Cutter Business Technology Journal* doesn't water down or delay its content with lengthy peer reviews. Written by internationally known thought leaders, academics, and practitioners — you can be certain you're getting the uncensored perspectives of global experts.

You'll benefit from strategic insight on how the latest movements in digital innovation and transformation, IoT, big data analytics and cloud, to name a few, are changing the business landscape for both new and established organizations and how cutting-edge approaches in technology leadership, enterprise agility, software engineering, and business architecture can help your organization optimize its performance and transition to these new business models.

As a subscriber, you'll also receive the *Cutter Business Technology Advisor* — a weekly bulletin featuring industry updates delivered straight to your inbox. Armed with expert insight, data, and advice, you'll be able to leverage the latest business management thinking to achieve your organization's goals.

No other journal brings together so many thought leaders or lets them speak so bluntly — bringing you frank, honest accounts of what works, what doesn't, and why. Subscribers have even referred to *Cutter Business Technology Journal* as a consultancy in print and likened each month's issue to the impassioned discussions they participate in at the end of a day at a conference!

Get the best in thought leadership and keep pace with the technologies and business models that will give you a competitive edge — subscribe to *Cutter Business Technology Journal* today!

Founding Editor: Ed Yourdon
Publisher: Karen Fine Coburn
Group Publisher: Christine Generali
Managing Editor: Cindy Swain
Copy Editors: Jennifer Flaxman, Tara Meads
Production Editor: Linda Dias
Client Services: service@cutter.com

Cutter Business Technology Journal® *is published 12 times a year by Cutter Information LLC, 37 Broadway, Suite 1, Arlington, MA 02474-5552, USA (Tel: +1 781 648 8700; Fax: +1 781 648 8707; Email: cbtjeditorial@cutter.com; Website: www.cutter.com; Twitter: @cuttertweets; Facebook: Cutter Consortium). ISSN: 2475-3718 (print); 2475-3742 (online).*

Subscription rates are US $485 a year in North America, US $585 elsewhere, payable to Cutter Information LLC. Reprints, bulk purchases, past issues, and multiple subscription and site license rates are available on request.

☐ Start my print subscription to *Cutter Business Technology Journal* ($485/year; US $585 outside North America).

Name | Title

Company Address

City | State/Province | ZIP/Postal Code

Email (Be sure to include for weekly *Cutter Business Technology Advisor*)

Fax to +1 781 648 8707, call +1 781 648 8700, or send email to service@cutter.com.
Mail to Cutter Consortium, 37 Broadway, Suite 1, Arlington, MA 02474-5552, USA.

Request Online License Subscription Rates

For subscription rates for online licenses, email or call: **sales@cutter.com** or **+1 781 648 8700.**

CUTTER CONSORTIUM
Access to the Experts

Opening Statement

by Barry Devlin

When writing a first draft of my book *Business unIntelligence* in 2011, I coined the term "biz-tech ecosystem,"[1] as follows:

> The evolving environment where business is fully or heavily dependent on information technology for all activities. IT adopts the role of partner and co-creator in business innovation. Business completely values and engages the full potential of IT.

Although the biz-tech ecosystem didn't resonate in the popular consciousness, I believe it better captures the essence of an emerging business environment than the competing — and now embedded — terms "digital business" and "digital transformation." Their shared focus on *digital* dangerously narrows the vision of those tasked with the data implementation to which digitization applies. Indeed, there's a lack of understanding of what might need to be implemented when it comes to both of these modern phrases. That a business should characterize itself simply as being digital feels more like analyst hand-waving than something that one could or should build. As for *digital transformation*, the absence of any indication of what must be transformed might pass for shoddy penmanship. It seems, therefore, that before discussing how to start building a digital business, we must first clarify what one is.

Defining Digital Business

The phrase "digital business," along with "e-business," dates to the last century — indeed, the last millennium — and the rise and fall of the dot.com era. By 2001, its meaning was already in need of clarification, as suggested by Adrian Slywotzky and David Morrison:

> … to think of digital business design as the sum total of the high-tech innovations multiplying around us is a fatally incomplete view…. The discipline of digital business design is not about wiring your workforce, converting your R&D to CAD/CAM, or selling products through your website. Instead, it is about using digital technologies to expand your company's strategic options. It is about serving customers, creating unique value propositions, leveraging talent, achieving order-of-magnitude improvements in productivity, and increasing and protecting profits.[2]

By 2016, in an issue of *Cutter Business Technology Journal* (*CBTJ*), Cutter Consortium Senior Consultant Borys Stokalski and Bogumil Kaminski described digital business transformation as:

> A journey of creating and combining specific business capabilities so that they give organizations a competitive advantage in the digital excellence domains in a way that reflects their chosen mix of strategic options. This journey is shaped also by the availability of critical resources — data, analytical skills, technology proficiency.[3]

This definition, and its focus on data and analytics, leads me back to my definition of the biz-tech ecosystem, which I propose as a synonym for a digital business:

> A digital business is a business that is wholly or heavily dependent on all forms of information technology in all activities planned and undertaken to achieve its strategic business goals. Using this technology, a digital business can harvest data from all sources, internal and external. It can contextualize and combine this data into useful information, from which valuable insights can be derived and informed actions can be taken at greater speed and frequency than possible in a traditional environment. Its processes, people, and organizations are optimized to take advantage of all the information available and to operate and evolve with the speed and efficiency thus possible.

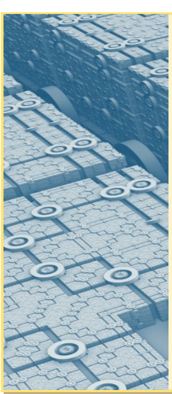
Where to Start Building a Digital Business

With this definition in hand, we are finally able to address the real question posed by this issue of *CBTJ*: what's the best way to start building a digital business?

Given that digital business affects all activities in support of every strategic goal, it is immediately clear that implementing a digital business implies an enterprise-wide effort. Furthermore, all three traditional classes of IT — operational, informational, and collaborative — are equally in the sights of digital transformation. In addition, today's Internet of Things (IoT) drives widespread and deep changes in operations, including extensive automation via artificial intelligence (AI) tools. Social media and highly mobile, tech-savvy customers are using new and extensive collaborative infrastructures to interact with businesses. Both these changes mandate significant new decision-making support using analytics, machine learning, and more. Enterprise-wide change of such scope is notoriously difficult to manage and succeeds only through a best-of-breed, cross-enterprise collaborative and change management effort.

Moreover, the old systems must be kept running and maintained even as IT turns toward new horizons. Of course, this is not a new problem. A variety of approaches have been proposed since the 1990s. These range from architectural (e.g., SOA) through methodological (e.g., Agile development) to organizational (e.g., bimodal IT). These seemingly distinct focus areas are not truly separate; they interact in often unforeseeable ways. More than 20 years' experience has proven time and again that, of these three aspects, it is the organizational challenge that often proves fatal to progress.

It is clear, therefore, that before we can address questions of a conceptual and logical architecture for digital business, we must first consider enterprise-wide methodological and organizational concerns. The typical IT focus on and rush toward new technology and tools as "magic bullets" targeted on digital transformation is disturbing. In short, building a digital business is a major undertaking, reminiscent of that old saying often applied to migrations of ERP systems: "It's like trying to change all four engines simultaneously on a 747 … in flight."



Upcoming Topics

The Next Wave of Cloud Computing
Frank Khan Sullivan

Fintech: Emerging Trends, Future Directions
Steve Andriole

Current experience building digital business is predictably disappointing. It is reported that 84% of digital transformation initiatives fail,[4] and that a large proportion of those failures can be traced to an unwillingness of organizations, both business and IT, to change behaviors. As Cutter Consortium Senior Consultant Greg Smith et al. recently pointed out: "A significant proportion of failures to realize the anticipated value from digital transformations can be directly traced to failure to appreciate and address human behavior associated with the transformation."[5]

How then to proceed?

Starting from the Data Warehouse

The fact that digital business begins with enormous amounts of diverse data and information offers a clue to a possible approach. Collecting and processing big data is something with which we are already familiar: could data lakes be a starting point? While attractive from a technology viewpoint — new, fast-evolving tools promise to eliminate old bottlenecks — data lakes still pose significant data management and governance challenges. The experience of many excavators of data lakes is that they often end up in data swamps!

One — and perhaps only one — concept in the IT world addresses the ingestion of large quantities of data/information from multiple sources, consolidating it for shared use, and delivering it for enterprise-wide and increasingly real-time use. The *data warehouse* is that approach. And it benefits from three decades of thinking, development, and successful implementation. I argue the case for the data warehouse under three headings: technology, information/data governance, and enterprise-wide focus.

Technology, as previously mentioned, is not the core issue in building a digital business. However, let's start by countering some key misconceptions about big data and data warehousing (DW).

Data warehouses traditionally handled volumes of data considered large for their time, until the advent of big data in the mid-2000s and more recently with the IoT. While the "3Vs" of big data (volume, variety, and velocity) still pose challenges to classic relational databases (RDBs) on which the data warehouse was based, advances in database technology and changes in DW practice mean these issues are now less of a concern than previously. Hadoop and NoSQL technologies in parallel processing and in-memory environments were developed in the last decade to counter the rigidity of structure and lack of scalability and performance of RDBs. Many of these techniques have since been embedded in RDBs; indeed, in many everyday big data circumstances across multiple industries, RDBs can support at least some of the core data processing and storage needs of a digital business.

Furthermore, most DW practitioners no longer believe that all data must be routed through a centralized enterprise data warehouse (EDW).[6] Multiple storage and processing technologies are included in the data warehouse and direct access to data sources — data virtualization — is now encouraged. The need to support real-time decisions means that the balance of power between database storage and processing engines is shifting.[7] In modern data warehouse architectures, the EDW is reserved for core business information that must be reconciled as a basis for enterprise-wide use.

This leads to the second aspect: comprehensive information/data governance. A digital business without it will fail and fail fast as unverified data is used for improper purposes, allowing multiple and inconsistent versions of truth about the business to proliferate. I use the term "information/data" rather than the more common "data governance" to emphasize the key role of the business meaning that must be assigned to data in a digital business. Data governance in traditional business — where most or all data is internally sourced and its meaning (mostly) is well defined — is challenging, and DW initiatives are often at the forefront of tackling these issues. The techniques and experience that DW teams have developed over the years will be invaluable in addressing the more complex and widespread challenges of governing data arriving at high speed and in high volumes from poorly understood or managed external sources and transforming it into useful, reliable information.

Governance addresses a fundamental issue: what is a true representation of the real world in the information systems of a business? As has become crystal clear recently surrounding the "fake news" dilemma, answering that question requires a human response, both individual and collective, and must be undertaken in a collaborative manner with collective shared interest and goodwill. Fortunately, within a business, we may expect that such virtues are more common than in politics.

Like information/data governance, enterprise-wide focus is an organizational rather than a technological issue. Successful DW program managers have long collaborated with all parts of the enterprise as they roll out a succession of projects. This alliance begins with the owners of data sources, who — even when internal — may set strict limits on what can be changed to meet new informational needs. The partnership extends to all potential users of warehoused information, who often do not know the use they will make of the information received until they see it. Thus, DW program managers become adept negotiators throughout the enterprise and at all levels of management to achieve their goals. The implementation of a digital business draws identical parallels.

> *The real value comes from the decades of experience in information/data governance and management, as well as the interpersonal and organizational skills that DW implementers have gathered.*

Starting from a data warehouse just makes sense. Of course, the architectural thinking and technology offer valuable intellectual capital to IT. But the real value comes from the decades of experience in information/data governance and management, as well as the interpersonal and organizational skills that DW implementers have gathered. As you will see from the articles in this issue, the contributors are on the same path.

In This Issue

In our first article, Daniel Power, an expert in decision support systems since the early 1980s, and Ciara Heavin argue that adapting to the changing business environment of a digital business is about much more than implementing new technologies like analytics, the IoT, and so on. Rather, business managers from the boardroom down must drive the adoption and skilled use of data-based decision making. Ultimately, Power and Heavin believe that middle managers are critical to digital business — making data-based decisions and selecting tasks that provide the necessary capabilities to deliver on a digital transformation vision and strategy. Organizational change management and skills development for these managers are thus at the core of digital business implementation.

Next, Martijn ten Napel, a practicing architect in the field of business intelligence (BI) since 1998, explores the challenge that has confronted him throughout his career: why do so many BI projects fail? His conclusion is that the struggle to achieve coherence between people, process, information, and technology has caused the complexity of the data landscape to grow out of control. His answer to the problem is the *connected architecture* — a framework and thought process for the organization of DW and BI projects. He believes it applies equally to digital business.

In our third piece, we hear from Michael Müller, an architect consultant in BI for almost 20 years. Like ten Napel, his view is that managing change is essential for a digital business initiative. This position emerges from his experience with BI projects where "Babylonian confusion" reigns, as business and IT lack a common vocabulary and an ability to communicate clearly about data needs and structures. Müller posits that digital business shares these same characteristics, but at a much larger scale because of the nature of big data. He offers an overview of a pathway for creating and maintaining conceptual-level data models that may be useful to implementers of digital business environments.

In the final article, Pat O'Sullivan, an IT architect and Senior Technical Staff Member at IBM, starts from the premise that the principles of standardization and conformity that were developed for the data warehouse are equally applicable to a digital business to deliver a consistent view of information to many lines of business. He explores the characteristics of a system of common metadata that can define the links between an existing data warehouse and an emerging digital business, describing the components and characteristics of this new metadata layer and how it is essential to fueling the growth of the AI capabilities of a digital business.

Conclusion

The contributors to this issue of *CBTJ* are of a single mind that a data warehouse is indeed a good starting point for building a digital business. They are equally consistent in their view that it is the organizational skills and information/data governance approaches, particularly with respect to modeling, that are most central to benefiting from a data warehouse implementation when defining and building a digital business. For experienced DW practitioners, these conclusions will hardly be surprising. However, given the more diverse backgrounds of implementers of digital business, there are important and valuable takeaways from the analysis of our contributors. My sincere thanks to them for their time and energy in sharing their valuable knowledge and experience.

Endnotes

[1]Devlin, Barry. *Business unIntelligence: Insight and Innovation Beyond Analytics and Big Data*. Technics Publications, 2013.

[2]Slywotzky, Adrian, and David Morrison. "Becoming a Digital Business: It's Not About Technology." *Strategy & Leadership*, Vol. 29, No. 2, 2001 (https://doi.org/10.1108/10878570110387671).

[3]Stokalski, Borys, and Bogumil Kaminski. "All the World's a Sound Stage: The Digital Transformation Journey in the Era of Hollywood Economics." *Cutter Business Technology Journal*, Vol. 29, No. 8/9, 2016 (https://www.cutter.com/article/all-world%E2%80%99s-sound-stage-digital-transformation-journey-era-hollywood-economics-492756).

[4]Rogers, Bruce. "Why 84% of Companies Fail at Digital Transformation." *Forbes*, 7 January 2016 (https://www.forbes.com/sites/brucerogers/2016/01/07/why-84-of-companies-fail-at-digital-transformation/#ad38090397bd).

[5]Smith, Greg, et al. "Putting the Human in Digital Transformation, Part I." Cutter Consortium Business Technology & Digital Transformation Strategies *Advisor*, 27 September 2018 (https://www.cutter.com/article/putting-human-digital-transformation-part-i-500696).

[6]Devlin, Barry. "Introducing the Production Analytic Platform." 9sight (blog), February 2018 (http://bit.ly/2Bt29Pi).

[7]Devlin, Barry. "Information- or Process-Centric? The Answer Lies Somewhere in Between." TDWI, 29 June 2018 (http://bit.ly/2uYYbJd).

Barry Devlin is a Senior Consultant with Cutter Consortium's Data Analytics & Digital Technologies practice. Dr. Devlin is a founder of the DW industry, having published the first architectural paper on the topic in IBM Systems Journal *in 1988. He is among the foremost authorities on BI and DW. With today's emphasis on big data, IoT, and the blending of the operational and informational worlds, Dr. Devlin is creating new methods and approaches to extend BI into true business insight.*

As an IT and management consultant, Dr. Devlin has guided BI and big data strategies and implementations in a wide range of business areas over the past 20 years. Clients include major banks, insurance companies, and retailers throughout Europe, the US, and South Africa. Dr. Devlin is a widely respected analyst, consultant, lecturer, and author of the seminal book Data Warehouse: From Architecture to Implementation. *His most recent book is* Business unIntelligence: Insight and Innovation Beyond Analytics and Big Data. *In addition, he is a regular contributor to the industry through blogs, journal articles, and Twitter. Dr. Devlin founded 9sight Consulting in 2008. Previously, he was an international consultant, manager, systems architect, software evangelist, and Distinguished Engineer with IBM for more than 20 years. He can be reached at bdevlin@cutter.com.*

FROM EVOLUTION TO REVOLUTION

Digital Business: Using Data in Decision Making

by Daniel J. Power and Ciara Heavin

Digital business is disrupting existing business models but adapting to a changing business environment is about much more than implementing new technologies like analytics, cognitive computing, the Internet of Things (IoT), or ambient intelligence. Digital business is a new, innovative way of doing business, and it goes beyond devising and transmitting a fresh corporate vision and strategy from the C-suite. Senior executives, including the CEO, CIO, and CFO, first and foremost must become data-based and data-informed decision makers. Even more important, there must be incentive for other managers in the hierarchy to follow their example. Ultimately, middle managers are responsible for making data-based decisions and selecting tasks that provide the necessary capabilities to deliver on a digital transformation vision and strategy.

Digital business is about understanding the value of data, using data, and enhancing its value using business analytics. A digital transformation strategy may implement a wide variety of digital business models. Thus, IT professionals must work with middle managers to determine how to integrate old and new technologies effectively in order to even begin to support transformation goals. Capturing, storing, and using data in a meaningful way is central to digital transformation and digital business.

This article briefly considers the need for new technologies to capture, store, process, and access new types of data. We reflect briefly on the technology evolution of data warehouses and data lakes as well as the evolution of computer-assisted decision making, including data exploration, analytics, and decision support. Arguably, the data warehouse still serves an important purpose and will continue to have a significant place in the IT infrastructure. It is important, however, to consider new technologies to support the need for real-time data, dynamic data warehouses, self-service business analytics, and algorithmic decision making. Finally, we argue that middle managers, in particular, must understand the importance of using data in facilitating digital transformation. They must become outstanding data-based decision makers and help transform their organization's decision-making culture.

Assessing Technology for a Modern Data Architecture

Data storage has evolved to handle new data sources, including very large volumes of streaming real-time data. Data warehouses to support decision making are not obsolete. Indeed, we may need to incorporate and integrate them into data lakes and cloud architectures. Thus, information technologists must learn to support big data analytics. For established organizations, the road to digital transformation is an uphill trek. There is — and will be — an ongoing and increasing need to provide analytics, decision support, decision automation, and transaction support using historical and current internal and external data in addition to data from newer digital business sources.

Data warehouses remain useful for organizing decision support data. Entities, the subjects in a data warehouse, may change and evolve and certainly more data about more attributes will be captured and stored. Metadata will become more important with managers preparing analyses. In general, involving more middle managers in data-based decision making is a key enabler of digital transformation and crucial in operating a digital business. Data exploration, decision support, and analytics are clearly central to data-based decision making. Therefore, managers in a digital business must have access to well-organized data to effectively find and analyze specific data to make effective, data-based decisions.

"Big data" is a colorful term that describes large sets of structured, semistructured, and unstructured data. These big data sets are continually increasing in size and may even grow too large for traditional storage and retrieval systems. Moreover, data may now be captured and analyzed as it is created and then stored. Big data is often described as high volume, high velocity, high variety, and high volatility. Decreasing storage cost and increasing use of machine data are major contributors to the big data phenomenon.[1]

New data sources present challenging problems and raise practical questions. Unfortunately, data is not integrated well in most organizations, data silos are still a problem, and a single version of the truth is still more a goal than a reality in most organizations.[2] IT managers must, therefore, develop a *sophisticated data storage strategy* that resolves questions of what data to store, where to store it, when to store it, how to store it, how to access it, who can access it, and other related questions.

Data Storage Technology Evolution

Everything that happens is being digitized. Customers are data, processes are data, supply chains are data, even employees are data. Indeed, all this digital disruption is changing how organizations and people behave. Until recently, data was stored for two primary purposes: processing transactions and providing decision support. Today, real-time data and other historical data are processed and stored for the purpose of automating decisions. Moreover, data is enhancing customer focus and support, is part of myriad new digital products, and is a means of identifying opportunities for new products and services.

> *Everything that happens is being digitized. Customers are data, processes are data, supply chains are data, even employees are data.*

A data warehouse is a database designed specifically to support decision making in organizations. It stores historical operating data used for reporting and for generating business intelligence (BI). Cutter Consortium Senior Consultant — and this issue's Guest Editor — Barry Devlin and his colleague Paul Murphy developed the "business data warehouse" concept in the mid-1980s.[3] By 1990, "the father of the data warehouse" Bill Inmon published his influential book *Building the Data Warehouse*, now in its fourth edition.[4] Creating a data warehouse is a data management strategy to collect, organize, store, and analyze large amounts of historical business data. There are two major approaches to storing data in a data warehouse: the dimensional approach and the normalized, relational approach.[5] A hybrid or combination approach is sometimes used.

In the *dimensional approach*, transaction data is partitioned into either "facts" (numeric transaction data) or "dimensions" (reference information that provides context).[6] With this approach, redundancy can be an issue. Inmon, a leading proponent of the top-down, *normalized approach* to data warehouse design, defined a data warehouse as a centralized repository for the entire enterprise.[7] The normalized approach looks to avoid redundancy. Today, data warehouses remain important and necessary for providing decision support and BI.

In a transaction processing environment, the assumption has traditionally been to store a transaction once and once only. However, decision support or data warehouse data storage is nonvolatile, and hence the assumption has been that multiple copies of the same data and information should be stored when the duplication improves query performance. Increasingly, real-time data is vital and indeed may have redundancy, but the data is unchanged and the redundancy may be important for analysis and for algorithmic decision making. Real-time data requires alternative storage to that used in a data warehouse or a transaction-processing system.

Here's where we come to the "data lake," a more recent buzzword in the analytics circle. A data lake is a metaphor for a large, unstructured set of data. In the physical world, a lake is a basin filled with water by rainfall, underground springs, and/or small streams. Similarly, a data lake is a storage structure for data fed by multiple sources. The data is diverse in terms of data sources and data types, and data is both structured and unstructured. Amazon Web Services (AWS) defines a data lake as "a centralized repository that allows you to store all your structured and unstructured data and at any scale."[8] Raw data is stored in a data lake in its native format directly from source systems. Techopedia explains that the "data lake architecture is a store-everything approach to big data. Data [is] not classified when [it is] stored in the repository, as the value of the data is not clear at the outset. As a result, data preparation is eliminated."[9] Surely, a data lake is more than a marketing term for Hadoop. The term is increasingly "accepted as a way to describe any large data pool in which the schema and data requirements are not defined until the data is queried."[10] A data lake may refer to the overall enterprise data architecture, or we can consider it as only a component.

Descriptive analytics, BI, and reporting remain important, and a modern data warehouse can support those important managerial needs. Supporting decision automation, as well as predictive, prescriptive, and diagnostic analytics, requires more modern data architectures. Therefore, we should revisit assumptions for data storage regularly. Data storage is no longer limited to implementing a relational transaction processing database management system or a static data warehouse. Often, data storage involves heterogeneous file structures and distributed processing. Certainly, we understand data storage better now because we have been doing it for 70 years, and data storage is easier and more robust because we have new technologies, including post-relational and mixed workload "translytical" databases in distributed off-premises computing environments.

Digital business is dependent upon digitization and digital data storage, but only data-based decision making can transform an organization and its culture so that data creates value for customers and stakeholders. The digital business must capture, analyze, and act upon internal and external data to enhance and support its mission. To help managers more effectively use data in decision making, IT leaders and experts should assess what is possible by investigating data and technology opportunities and the associated costs. Digital business means the organization, on the whole, has a culture that tries new technologies and assesses what works.

Evolution of Computer Decision Support

Decision making in business has evolved from a largely informal, intuitive process for structured and unstructured decision situations to a much more systematic process, especially for routine, structured decisions. Today, analytics and computerized decision support can help managers make better choices in semistructured and even unstructured situations. Therefore, managers should be curious and seek evidence and answers from data for previously unasked and/or unanswered questions. Indeed, they should become *data explorers* and sophisticated decision support users.

Data exploration is a process in which nontechnical managers and staff use data visualization and menu-driven query tools for one-time or ad hoc analysis and search. Some data exploration may be part of a more systematic analysis. This step is sometimes termed "descriptive" or "exploratory" analysis. Managers who interactively explore data sets suspect that a relationship exists and try to confirm or deny that suspicion. The informal nature of data exploration that involves testing inferences can be concerning to those who focus on more formal survey research. The mere act of exploring can, however, generate more testable hypotheses. The issue, therefore, is more about when the proof is adequate enough to act on regarding an intuitively derived relationship. Managers should routinely ask for facts and data.

Only data-based decision making can transform an organization and its culture so that data creates value for customers and stakeholders.

Data exploration should begin with understanding a data set and checking to see if it is relevant and useful. A manager should first ask, "Does the data set contain the variables of interest?" If the goal is predicting sales, another thing to ask would be: "Is the sales data usable, and are there possible predictor variables in the data set?" Other important questions include: "Is the data set high quality?" and "Is the data set clean and are data types compatible?" If the answers are "yes," then managers should use self-service tools to analyze and make more sense of the data.

It is also important to consider the ethical implications of data exploration. For example, does the data set contain sensitive, identifying variables? Do you have the appropriate permission to analyze and use this data? Organizational policies and legislation should guide the behavior of managers on these and related topics. It is essential to encourage and promote open discussion and proactive behavior to ensure the ethical use and construction of analytics and computerized decision support.

Data exploration has a critical, possibly even indispensable, role in most analytics tasks and projects.[11] Managers must learn to explore data; that is, become comfortable trying to figure out "what the data says" about business questions. Managers should challenge the data and pursue answers to business questions using data and technology. Modern managers should be both inquisitive and persistent in seeking facts. They should then confirm the findings and, finally, share those discoveries with others.

There are four broad categories of business analytics:[12]

1. **Descriptive/business reporting.** What is happening? What happened?
2. **Diagnostic.** Why did it happen?
3. **Predictive.** What will happen and when? What will happen next?
4. **Prescriptive.** What should happen next?

These categories are common to business analytics and data analytics practitioners. We can perform each using many data sources as input, including big data sources such as machine data and social media. *Descriptive*, or reporting, analytics describes or summarizes past results, actions, or activities. (The term "business intelligence" encompasses descriptive analytics.) This type of analytics primarily uses data aggregation and statistical tools, such as averages and differences. *Diagnostic* analytics emphasizes understanding the causes and why something happened. This category uses drill-down tools, interactive data visualization, and data mining. *Predictive* analytics extrapolates data about the past into the future and potentially helps decision makers understand the future. This type of analytics uses more complex statistical models (e.g., regression) and correlation and forecasting techniques (e.g., moving averages). Finally, *prescriptive* analytics tools help quantify the impact of a decision before making one and prescribe an action to take based on analysis of outcomes. This category uses various tools, including optimization, simulation, scenario analysis, and case-based reasoning.

All four types of analytics can be embedded in specific decision support systems (DSSs). For example, descriptive analytics is often part of data-driven DSSs providing BI. Diagnostic analytics can be a major component of some knowledge-driven DSSs. Predictive analytics, especially using forecasting models, is part of some specialized model-driven DSSs. Finally, organizations can incorporate prescriptive analytics using quantitative models or knowledge rules in DSSs for repeated use.

Data-based decision making refers to an ongoing process of collecting and analyzing different types of data to aid in decision making.[13] A skilled, data-based decision maker follows a process that begins with asking the right questions then answering them using facts, relevant data, and analyses prior to making decisions. "Data-informed decision making" is a phrase used when data and facts are an influential factor in decision making, but not the only factor.

In general, the more a manager knows about an organization and its data — and the more comfortable a manager is with analysis tools — the more valuable the exploration of a data set. When managers can fully explore and test/refine their hypotheses using a cleaned and transformed data set, then more complex analyses are possible.

Summary and Conclusions

Analysis of data with business analytics is an integral component of a successful organizational digital transformation strategy.[14] Business analytics refers to a set of innovative tools and techniques used to generate insights from large, diverse data sets. The use of analytics can offer individuals, organizations, governments, and the global society "data-based" perspectives on existing challenges and possible solutions. Analysis can potentially provide facts to improve data-based decision making. Business and data analytics help find meaning in an increasingly complex environment.

Data — even "good data" — on its own will not result in good or better decision making.[15] There are major challenges with succeeding in digital business transformation, including the biases and habits of managers, lack of training, resistance to change, limited analytical skills, and cultural factors and constraints. Indeed, the benefits of analytics and decision support are always constrained or limited by the manager (or managers) who use the results to make decisions. To achieve meaningful integration between business analytics tools and technologies, managers must use analytics in parallel with a data-informed program for change.

It is difficult to negotiate the balance between an organization's approach to analytics and a data-based management strategy. To achieve this dual focus, it is essential that managers meaningfully engage with the process of analyzing data to leverage new data insights and integrate them into the organization's decision-making processes. Shvetank Shah, executive director of the Corporate Executive Board, and his colleagues suggest managers become better trained to use new analytics tools, paying particular attention to building analytics into managerial decision-making processes.[16] If achieving balance is difficult, managers run the risk of investing in new, expensive data analytics technologies that staff ultimately will not use. Analytics pioneer Tom Davenport noted 10 years ago that “we’ll have to have more and more analytical people in organizations to make effective use”[17] of analytical capabilities.

In this era of digital disruption, more employees need analytical skills, especially managers. Good quality data and information resources must be easy to locate and managers must have access to data relevant to their job for self-service analytics. Analyzing and using data are the responsibilities of everyone in an organization, especially C-suite executives and middle-level managers. Managers must become knowledgeable, independent data users and data-informed skeptics. Data varies in value, quality, and reliability, and managers must become intelligent data consumers and analysts to build a modern digital business and to create business value. Using data and analyzing data are every manager’s job.

Finally, managers should implement the “right” amount of analytics to support a data-based culture. Such an organizational change should lead to better data-based decision making and data-based management. The overriding transformation goal is to use business analytics and computerized decision support to assist in managing directed technology change. This will ultimately lead to a transformed business model and competitive advantage.

Endnotes

[1]Power, Daniel. J. “Using ‘Big Data’ for Analytics and Decision Support.” *Journal of Decision Systems*, Vol. 23, No. 2, 2014 (https://www.tandfonline.com/doi/abs/10.1080/12460125.2014.888848).

[2]”Survey Reveals Most Organizations Struggle to Extract Value from Operational Data.” Press release, Devo Technology, 8 August 2018 (http://dssresources.com/news/5013.php).

[3]Devlin, Barry, and Paul T. Murphy. “An Architecture for a Business and Information System.” *IBM Systems Journal*, Vol. 27, No. 1, 1988 (https://pdfs.semanticscholar.org/c22c/e1eeafb01f0682e194a2a22349aa141b78f6.pdf).

[4]Inmon, William H. *Building the Data Warehouse*. 4th edition. Wiley, 2005.

[5]George, Sansu. “Inmon or Kimball: Which Approach Is Suitable for Your Data Warehouse?” *ComputerWeekly*, April 2012 (https://www.computerweekly.com/tip/Inmon-or-Kimball-Which-approach-is-suitable-for-your-data-warehouse).

[6]Kimball, Ralph. *The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses*. Wiley, 1996.

[7]Inmon (see 4).

[8]”What Is a Data Lake?” Amazon Web Services (AWS) (https://aws.amazon.com/big-data/datalakes-and-analytics/what-is-a-data-lake/).

[9]”Data Lake.” Techopedia (https://www.techopedia.com/definition/30172/data-lake).

[10]Rouse, Margaret. “Data Lake.” TechTarget, May 2015 (https://searchaws.techtarget.com/definition/data-lake).

[11]Russom, Philip. “Data Exploration in the Age of Big Data.” Transforming Data with Intelligence (TDWI), 19 November 2013 (https://tdwi.org/articles/2013/11/19/big-data-exploration.aspx).

[12]Power, Daniel J., and Ciara Heavin. *Decision Support, Analytics, and Business Intelligence*. 3rd edition. Business Expert Press, 2017.

[13]Power, Daniel J., and Ciara Heavin. *Data-Based Decision Making and Digital Transformation*. Business Expert Press, 2018.

[14]Power and Heavin (see 13).

[15]Shah, Shvetank, Andrew Horne, and Jaime Capella. “Good Data Won’t Guarantee You Good Decisions.” *Harvard Business Review*, April 2012 (https://hbr.org/2012/04/good-data-wont-guarantee-good-decisions).

[16]Shah et al. (see 15).

[17]Power, Daniel J. “Tom Davenport Interview: Competing on Analytics.” DSSResources.COM, 27 May 2007 (http://dssresources.com/interviews/davenport/davenport05272007.html).

Daniel J. Power is a Professor of Information Systems and Strategy at the College of Business Administration, University of Northern Iowa. He is also the Editor of DSSResources.COM. Dr. Power has published more than 50 refereed journal articles and book chapters. His articles have appeared in leading academic journals, including Decision Support Systems, Decision Sciences, *and* Journal of Decision Systems. *Dr. Power has authored or coauthored six books on computerized decision support. His two most recent books, coauthored with Ciara Heavin, are* Decision Support, Analytics, and Business Intelligence *and* Data-Based Decision Making and Digital Transformation. *Dr. Power earned a PhD in business administration from the University of Wisconsin-Madison. He can be reached at daniel.power@uni.edu.*

Ciara Heavin is a Lecturer/Researcher in Business Information Systems at Cork University Business School, University College Cork, Ireland. Her research focuses on opportunities for using information systems in the global healthcare ecosystem and in digital transformation. As Co-Director of the Health Information Systems Research Centre (HIRSC), Dr. Heavin has directed funded research in the investigation, development, and implementation of innovative technology solutions in the health/healthcare domain. She has published articles in several top international IS journals and conference proceedings. Dr. Heavin coauthored two books with Daniel J. Power: Decision Support, Analytics, and Business Intelligence *and* Data-Based Decision Making and Digital Transformation. *She earned a PhD from University College Cork. She can be reached at C.Heavin@ucc.ie.*

Taming the Data Beast: Inheriting Key Insights from Data Warehousing Practice

by Martijn ten Napel

Digital transformation is not so much about the use of technology. Instead, it concerns creating fluidity in what used to be demarcated boundaries between stakeholders in a business process. Digital information flows enable fluidity. Direct communication with consumers along with a quick adaption to their shifts in taste — or fine-tuning to individual preferences — are changing the way businesses operate.

Technology has enabled us to build systems that support business processes across the boundaries of physical and legal entities. Technology has resulted in a restructuring of market conditions. Consequently, the competitive space has been altered and businesses are adjusting to this change.

> *Putting information to work for your business is a collaborative effort ... a continuous process, not a project with a finite result.*

But adjusting isn't easy, and information is not a fixed stream. The information itself and where it originates are dynamic. The process of pulling it together, deriving insights from it, and acting upon those insights is quite a challenge. Attempting to achieve the desired business agility — an attempt that will never reach a final state — is hard work.

As a result, businesses need to significantly adapt the way they govern their information landscape. The old demand-supply model of IT governance isn't adequate; we need a governance model that accommodates extreme fluidity.

Riding the Data Beast

Agility is one of the key objectives in a governance model that accommodates fluidity. Agility is a result of the culture of an organization, not a state that can be achieved by implementing the rituals of an Agile working method. The foundation for this statement comes from the observations I have made and the experiences I have gathered over 20 years across a diverse set of businesses. It is difficult to reach a state of agility that is sufficient. There are many reasons why this is difficult, but I observe one common theme: universally, people struggle with the data beast. Putting information to work in a way that is productive is like riding a rodeo bull, whether you call it business intelligence (BI), business analytics, data warehousing (DW), big data, information management, data governance, or digital transformation.

Putting information to work for your business is a collaborative effort; an effort that is a continuous process, not a project with a finite result. Every time you feel you have obtained a definitive result, the data beast starts to buck again.

So what is this beast and why does it buck?

What Is the Data Beast?

The collaborative effort is a consensus model. You and other stakeholders need to discuss and arrive at a shared conclusion of what information means to you and how to adjust a business process. The adjustment can change the way a business process runs or lead to process redesign in digital businesses.

Information gathering and recording are plagued by fragmentation, context switching, and volatility. These problems seem to be inherent to working with data and constitute the data beast. The contradiction between, on the one hand, the search for consensus and, on the other, the fragmentation, context switching, and volatility of information that dilute this effort (see Figure 1) is a never-ending rodeo ride.

The reasons behind fragmentation, context switching, and volatility are misunderstood. This triple plague is often seen as either an imperfection that requires fixing or a roadblock to digitization. As a result, the way we deal with it is often counterproductive.

Before diving into this problem and suggesting how you can rethink the challenge of digitization, let me first explain what I mean by fragmentation, volatility of information, and context switching.

The Nature of the Beast

Fragmentation

To get the whole picture of the business, information must be assembled from different software systems or data stores. This is fragmentation. In DW or enterprise application integration, this phenomenon leads to the need for integration.

With digitization, information is becoming increasingly fragmented. Connectivity, the driver of digitization, enables us to compose business processes across supply chains and across cloud services. Information is not integrated but rather connected across the steps in the business process or the particular services that deliver one or more functions of the business process. For instance, Salesforce.com can be used as a service within different business processes. It enables the recording and retrieving of customer interactions across processes or across the different communications channels and social media outlets that engage customers, but its data set is isolated from the data sets in other parts of the business process. Consequently, keeping the definition of information in sync is a challenge.

Disparity of data is not the only consequence of using a service. Loss of ownership has an even more profound impact. For instance, think about the message service WhatsApp. For many businesses, it plays an important role in customer interaction, but its functionality is determined by Facebook, not by the business using it. If a business composes a business process from different services it does not control, fragmentation makes it harder to keep information consistent throughout the business process as opposed to when a process is composed from systems the business controls.

Volatility

The volatility in the information landscape is increasing because businesses are interacting and assembling their processes in different ways; for example, by enabling and switching to cloud services. Cloud services offer

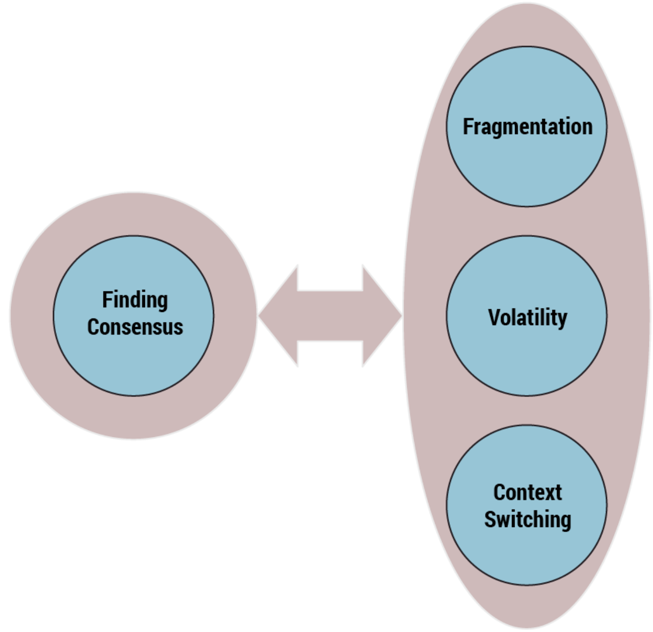


Figure 1 — The core components beneath the information use process.

new functionality at an ever-higher velocity to remain competitive. Maintaining information consistency is hard work when the landscape in which data is recorded changes all the time.

Since services are for the most part no longer owned by the businesses that use them, businesses must keep up with the velocity set by the service provider. Many companies struggle with this externally driven change in IT operations. Businesses cannot plan for this change; they must simply keep up. People using and retrieving information from these services are no longer in the driver's seat but are merely passengers along for the ride.

Fragmentation and volatility are not exclusive to DW practice, but data warehouses are often the systems where the volatility across all systems converges and where fragmentation has a significant impact on being able to keep information consistent over time. The BI community has learned how to deal with this, and those methods can be applied to the digitization of business processes.

Context Switching

Context switching occurs when information is used outside the context in which it was recorded. It is what happens when you ask the question, "What does this piece of information mean?" A simple example can be found in a financial transaction that leads to an entry in a financial ledger designed to record the transaction. When you aggregate financial information across ledger entries and ask the question, "Are we on track with our sales goals?" you interpret the ledger entry in another context. Depending on the answer, you might follow different courses of action.

Context switching is the least understood and hardest part of working with information. It is where consensus on how to define a shared context for use of information must be found. What makes DW difficult is that you must embed this consensus on meaning, in the context of a particular use, in your data models.

Most DW practitioners have rejected the adage "one version of the truth" and have replaced it with "one version of the facts; multiple views on the facts." It is an acknowledgment that the same data point, recorded for a specific purpose, can have different meanings in different business contexts.

People miss information that was disposable in the customized solution they use, and they try to find other ways to get to that information.

Context switching is a phenomenon that used to be confined to data modeling in the DW and BI functions. Connectivity, together with the low cost and massive growth of compute power, has made it possible to integrate the decision-making process directly into business processes. As such, the challenge of context switching is dispersed across the whole business and is no longer an exclusive BI or analytics phenomenon. Dealing with multiple contexts of use is another lesson that can be carried over from DW to the digitization of business processes.

Context switching is where the data beast is really trying to throw you off its back. With the digitization of business processes and the fluidity in the boundaries of a business, the number of contexts in which information is interpreted has grown enormously. When finding consensus, or automating decisions within a digitally enhanced business process, you must be very careful to understand the context of the participating actors.

And here is the tricky part: consensus on "What does this piece of information mean?" will often lead to a *next action*. That next action might lead to a change in the business process or in the context of the next question (a new or adjusted interpretation). BI professionals know that people can only tell you what information they need once they have seen it. Insight gathering and achieving consensus make for an interactive and progressive change process that often leads to more volatility and, if you are not careful, increased fragmentation.

Taming the Data Beast

How to deal with these phenomena — fragmentation, volatility, and context switching — is where opinions start to diverge. The most common approach is to strive for unification and to create standards and standard services.

Most people agree that volatility is a given. Many architects will reason that defining standard components or standard building blocks is a precondition for achieving agility in composing or changing business processes. The line of thought is that standardized components can be swapped in and out when necessary, with a short time to market, low cost, and limited impact on other standardized components. This is the engineering point of view, comparable to a builder using standardized components in different house designs that can be chosen from a catalog.

I've been advocating this approach for a long time but have never arrived at a satisfactory state where this standardization led to a smoothly running and agile information stream. On the contrary, the more volatility popped up and the more "standard" solutions were introduced, the harder it became to reach consensus and deliver fitting information solutions. What I've noticed is that using standardized building blocks in IT architecture always increases information demand. People miss information that was disposable in the customized solution they use, and they try to find other ways to get to that information. Part of the demand is the human aversion to change and projects not taking care of educating the users, but part of it is also IT people making decisions on priorities without fully understanding the business context of users. With increased volatility, the information demand rises beyond what a company can cope with and people start to improvise and create local solutions. As a result, fragmentation increases exponentially.

In my DW and architecture practices, I've tried different ways to approach a solution, taking what I've learned from other projects. Moreover, I've been a proponent of using Agile working methods to deal with the learning curve resulting from context switching. Such methods enable us to take steps forward in maturing the process of working with information.

Maturity is the result of two processes being managed in an appropriate fashion. One is the collaboration process between information users and information provisioners, to better understand the business context of users, but also for users of information to understand what data can and cannot do for them. The other process is the prioritization of new information requests based on business value. This requires users to collaborate, to see beyond their own business context, and to be able to decide what is of more value to the entire organization. Agile working methods work, and when implemented the right way, I've seen consistent improvement. But it is not enough.

Time and time again, fragmentation and volatility complicate this process. Reaching higher maturity levels in creating consensus out of information and in acting on this consensus is challenging. Often, businesses slide back after a while until a new manager tries again and initiates a new project or a new change. This has been puzzling me for a long time. We know the problem, and we brought along our data architectures and our ways of working as a BI community to offer solutions. Why are we failing to make our practices stick?

Changing Our Approach to the Data Beast

What We Can Learn from DW Practice

I believe we have been denying the true nature of the data beast. I am convinced that we have misunderstood the causes of fragmentation and volatility. We have regarded them as imperfections that we have to battle, instead of as artifacts of our own behavior.

Put DW architects together in a room and the discussion turns quickly to the best way to deal with both fragmentation, which leads to integration challenges, and volatility in the information landscape over time. Now, many architectures are available, but there is a tradeoff around when to integrate and what it means for the changeability of information. The common complaint I've heard in 30 years of DW has been "Why does creating, extending, or changing the data warehouse take so long?"

The answer is that dealing with fragmentation, volatility, and context switching is hard work and it needs close collaboration with the users of the information. Fragmentation and volatility drive the need for a consensus-achieving process and are, at the same time, the results of the consensus-achieving process. Using information and achieving consensus on its significance through discussion leads to a next action. The results of that action could well be a change to the business process and the supporting IT systems. It could even lead to a redesign, where parts of the process are reimplemented in services or new parts are added to the business process, sourced by a new supporting IT system. In turn, changed or additional information needs to be collected from these new or changed systems or services. Context switching complicates our understanding of, and our reaction to, dealing with fragmentation and volatility. Test, experience, observe, debate, and adjust are how we humans progress in business. With the digitization of businesses, this challenge intensifies as volatility increases.

In digital businesses, fragmentation is a design decision to deal with the fluidity of the business processes and business boundaries. Understanding the consequences in terms of the energy it takes to keep information consistent across a fragmented data landscape is among the key DW insights to be carried over.

Reaching higher maturity levels in creating consensus out of information and in acting on this consensus is challenging.

Dealing with the Data Beast

Once you can accept that the beastly nature of working with information is created by your own actions and is an inherent part of the collaboration process that makes information work for you, you can finally start to formulate solutions.

One lesson to take from DW practice is to never assume a future state. The information landscape will always be fluid. Volatility is a given. At the same time, we must deal with decisions made in the past: historical patterns in data cannot be undone. As a discipline, we have developed coping mechanisms.

Architects have formulated concepts like loosely coupled services to create connectivity between the components of a business process and its master data; that is, to integrate the components into a coherent process. The scale at which loosely coupled services are deployed has changed, and the ownership model certainly has changed with digital transformation.

Fragmentation isn't bad; it can be used to create the conditions to achieve agility in the data landscape if it is used as a design principle. It allows you to string together software services or components. Volatility and fragmentation across different software components are expressed in fragmentation of master data. Acquisitions and mergers, partial replacement of ERP systems for cloud services, and changes to business processes result in fragmentation of master data.

When we talk about context switching and a need to reach consensus, master data is the integration point. Fragmentation makes integration harder. On the other hand, having multiple expressions of the same master data item is not an unfortunate side effect of context switching, but the result of having to deal with multiple contexts. Multiple expressions of the same master data item might even be inevitable with the increasing fuzziness of business boundaries.

My experience at Free Frogs, a cooperative BI company, has shown that applying the principles of "loosely coupled" to master data and containing fragmentation within a framework that governs the collaboration process will lead to the design patterns of solutions that fit this collaboration process. This is *designed fragmentation*, or "connected architecture," as dubbed by Free Frogs (see Figure 2). This framework is a thought process more than a recipe.

Connected Architecture in Designing the Arena for the Data Beast

Human interaction is key. This is the first line in Figure 2. Working with information requires a collaborative process involving all disciplines in order to facilitate reaching consensus on the meaning of information in its various contexts and to determine the actions to be taken. In other words, context switching is a learning process that needs close collaboration among three groups of people with different skills and different responsibilities: those who use information and create value out of it, those who have the analytical skills to provision and interpret the information for users, and those who have the skills to prepare the information and who can automate the data streams involved. Close collaboration needs reinforcement into solutions to align the business process to the consensus.

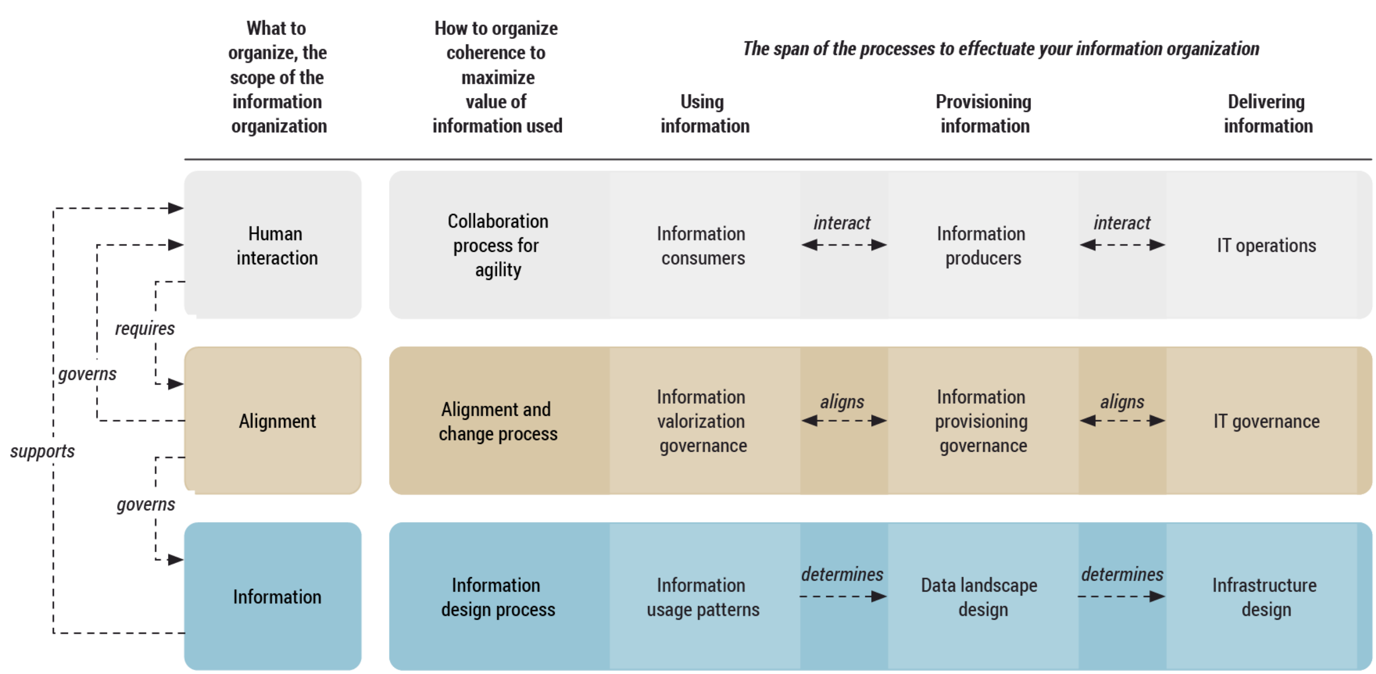


Figure 2 — The connected architecture framework.

The process of alignment, the second line in Figure 2, requires an architecture process that reevaluates the design continuously and adapts to the volatility, thus enabling agility. Architecture cannot create agility, but architecture can create the conditions for agility to thrive.

You need to focus on what brings value. Most information has limited value, but you need the process of finding consensus to determine which information is valuable. Of course, what is valuable changes over time, and volatility also changes the value of information. You need a guided process to focus the effort of collaboration on what has value. It won't happen by itself; you must organize the collaboration process. The architecture process must find a delicate balance between agility and alignment.

This implies that the architecture process shapes the information design process — the third line in Figure 2 — which uses fragmentation as a design principle to safeguard the required agility. Where this agility is needed emerges from the collaboration process. We know these needs are not stable by the nature of the collaboration process.

This task of finding balance is not simple. An architect cannot get it right all the time. Like all people involved in the process, you learn from the collaboration, from inspecting the results, and you build and change both the landscape and the conditions for the human interaction process. Your design process must be incremental, and your data landscape design should be able to facilitate the evolution of insight. This evolution originates in the outcomes of context switching.

Once you accept that fragmentation, volatility, and context switching aren't disruptors to digitizing your business processes but instead are ingredients to create solutions that support them, it becomes much easier to wrap your head around what is necessary.

This way of thinking originates in dealing with the challenges caused by fluidity in the information that flows to and from a data warehouse. These challenges are not unique to DW and the framework can be applied to other parts of the digital business as well.

The governance model needed to accommodate fluidity is challenged by the nature of working with information. Once you have more knowledge of what this nature is, you can adapt your information governance model to work with that nature instead of trying to battle it. It is impossible to deliver guidelines on how to do this in the scope of a single article, but you can take a second look at your digital transformation efforts with the following three takeaways in mind:

1. **Take a critical look at who is represented in your digital transformation teams.** Is the mix of users, business process domain experts, information professionals, and IT professionals balanced? Do they collaborate on equal terms?

2. **Hire a few architects who understand the nature of working with information and are equipped to guide the teams.** They should blueprint your digital transition, paying careful attention to the balance of the level of fragmentation in the data landscape, but leaving the details up to the teams.

3. **Incorporate the professionals of your DW or BI departments into your digital transformation teams**. This helps to define integration patterns and deal with different expressions of the same master data item across different systems.

Martijn ten Napel is an experienced architect in the data domain. He is currently working on ways to provide organizations with the tools needed to help them understand the consequences of their behavior and to simplify their data landscape. Mr. ten Napel publishes on this subject at preachwhatyoupractice.nl. He earned a master's of science degree in economics and business administration from Erasmus University Rotterdam, the Netherlands. He can be reached at martijn.ten.napel@freefrogs.nl.

It's All About That Data

by Michael Müller

A digital business thrives on its data. The business must create value from the data not only to gain success, but also to close the gap with successful competitors — especially with new competitors that have had the opportunity to build a greenfield solution[1] and utilize their data more extensively to create increased revenue and business value.

A less expensive way than a greenfield approach to achieve more customer value is by building new processes or using existing data from the data warehouse in new, innovative ways. Unfortunately, there is often very little knowledge about the underlying data. Although business users tend to have a clear picture about the information stored in the data, they are often unable to express its precise meaning in a way that it is understandable to IT or other departments. IT, on the flip side, is often unable to translate its data concepts in a way that is understandable to business users. This situation breeds a great deal of miscommunication and misunderstanding. Thus, we need a common vocabulary to end the pattern of discussions that result in Babylonian confusion.

This Babylonian confusion is well known to anybody implementing or developing a data warehouse. In a data warehouse, a conceptual data model offers the opportunity to overcome the lack of a common vocabulary. A high-level conceptual data model delivers a common language for describing the lifecycle of data. With this model, each interaction of the customer journey can be seen along the data trail, delivering a deeper understanding of how the data moves through IT systems. There's a greater focus on the big picture, describing the necessary results and artifacts rather than giving step-by-step instructions of how tasks and processes should be done.

This big picture approach makes clear the relevant information and necessary data. A conceptual data model helps declare the requirements for a business intelligence (BI) system and the requirements of the IT systems. This model allows the team to analyze strengths and weaknesses of various possible solutions. These solutions can be tested before being implemented. With a clear map of the company data, it becomes easier to find a way to integrate external data and easier to spot whether the external data generates revenue and business value.

A conceptual data model shifts the focus to valuable metadata. In the past, conceptual data models were often costly and poorly maintained. The effort involved in keeping the model up to date was simply too high. To achieve success with a conceptual data model, we must cut away the excess and integrate the model with the existing metadata. Harvesting the metadata from existing systems saves time. Constant checking of the model and the corresponding systems is required to ensure future accuracy.

Managing Change Is Essential

Digital business means changing the current business model by delivering customer value through a new combination of the digital and physical worlds. This novel way of creating customer value is done with data (i.e., machine-processed information), connecting the physical objects in the same way as power lines. But unlike a power grid, the connection through data is often not well documented. Since data is not a physical object, people's understanding of its purpose and content can vary widely. Therefore, a greenfield approach looks very promising because all participants typically have the same clear picture of the result. Planning is done in parallel across all areas to yield the best solution with a clear architecture. And on top of all that, the project has the full attention of high-level management — something that the cost alone will guarantee.

Now, everything may look positive on the surface, at least until the change necessitated by the solution becomes evident. Then there is a high risk that the needed change is not compatible with the existing solution or, worse, with the existing understanding of the business. Thus, it is not only the solution that must be maintained over time; an understanding of the underlying picture by all must be ensured as well.

In the end, attempting a greenfield approach comes with the same problems as an evolutionary approach. Throughout evolutionary change, a common picture must also be maintained. This problem has been at the root of BI systems for several years now. Every five to seven years, someone tries a greenfield approach in BI. But only a handful of companies have managed to maintain and extend a solution for more than 10 years — those who have found a common picture. These new, successful BI systems started with a common picture and a common language. One of the essential success factors is avoiding fruitless general discussion and instead focusing on facts and results, while allowing differences of opinion. Because these differences reveal a lot about "what is really going on," having them out in the open helps change perspectives.

A Data Warehouse Reveals a Common Picture

In data warehouse development, a glossary addresses the need for a common language. This glossary has evolved in recent years into a conceptual data model. A conceptual data model defines business objects in the same way as a glossary but has the advantage of also showing the relationships between business objects. A conceptual model is a high-level model that allows all IT systems to redefine their part of the conceptual model by creating their own logical/physical data models. The design of a data warehouse usually calls for the integration of the existing IT systems' data in such a high-level conceptual data model. The resulting model is then used to generate a first version of a logical data model for the data warehouse. The generation of this first version is a huge part of data warehouse automation because some part of data warehouse development is done automatically.

A Conceptual Data Model Provides the Vocabulary

A conceptual data model is a very good starting point for implementing a digital business. In such a model, there are business objects like "customer" or "product" along with dependent business objects like "order" (see Figure 1). Dependent business objects are usually the result of a business transaction; they link the content of the transaction to the assets (business objects). Business users easily understand this logic. A conceptual model is an information model — a model made for human consumption rather than for computers but still formal enough that computer-readable data can come out of it.

An individual instance of a business object is identified by a business key (also shown in Figure 1). Business keys are easily found by asking business users, "How do you refer to a specific customer/order/product?" Technical keys — keys used in the IT system for identification — usually exist as well. There might be more than one technical key because a particular business object is used in several IT systems. Put all the keys into the model because they may provide insight into future problems that might need solving at some point. Gathering all this information is part of getting to know "what is really going on."

Ultimately, there may be problems surrounding definition. For example, sales may define a customer as someone who has placed an order in the last six months, while marketing may say everybody interested in our products is a customer. The resolution is to show the difference by using precise definitions, such as:

- A *marketing customer* is somebody in our customer database.
- A *sales customer* is somebody in our database who has placed an order in the last six months.

Honor the Differences Between Departments

Such precise definitions avoid long discussions that attempt to come up with one all-encompassing definition and allow differences and various focuses to become clear. This is good; different departments have

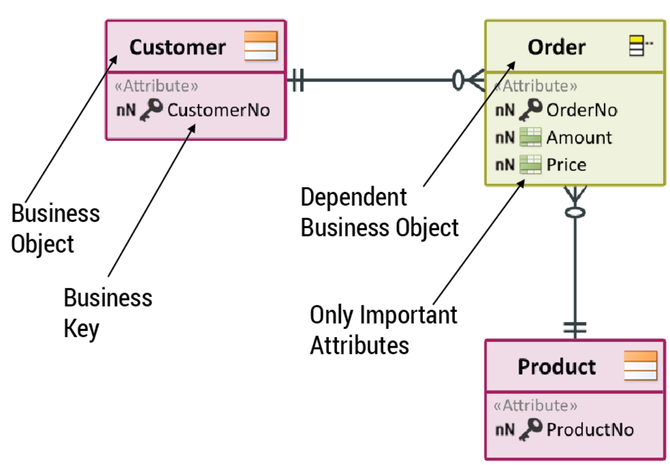


Figure 1 — A simplified conceptual data model of an order.

different jobs to do. Appreciate business users by honoring their view on their own area of expertise. And, to return to our customer example, if customer care has a third definition of customer, that's fine, too. Everybody now has the vocabulary in hand to define the differences. Eventually, these differing definitions will bring some inherent problems out into the open and might help close age-old trenches between departments. At the very least, senior management of these various departments — when identifying a need for change arising out of differing definitions — will know what to change.

Ensuring that the conceptual data model stays up to date is a very good way to make sure future development will not go astray.

Moreover, these definitions automatically deliver the important attributes to a conceptual data model. They will be needed in the BI system to make proper reporting possible for different departments on, for example, the number of marketing customers, the number of sales customers, or the number of support customers. Laying out the differences and providing everyone with the same understanding as well as a vocabulary to address the differences clears the way for making and implementing strategic decisions. Without a common understanding, the results are often disastrous because people feel threatened and only work with their current definitions. Proper reporting when a common understanding is lacking requires people to find their own precise definitions, a process that takes time — time that is usually not available. If differing definitions are accepted and people understand the contrasts, they usually find a good way to integrate the definitions into their thinking and accept and work with them. Because they see the benefits, the change becomes something they desire and will find the time to implement.

Harvest Existing Knowledge

The discussion of the conceptual data model cannot start with the open-ended question, "What are your business objects?" That kind of approach usually results in the business asking the typical question, "Well, what business objects are there?" — meaning that business users prefer to choose from available data rather than pick just one. To avoid this standoff, IT must prepare by understanding which business objects might be available and by using the language of the business. The knowledge about the business objects is there, if one knows where to look. Before going in depth with the business, look at the important existing systems. Look at existing documentation or old requirements documents, but don't take them at face value. Remember that implementation usually brings in a new perspective, change that is seldom documented after the fact. Instead, harvest the metadata from the existing database, creating a data model, and collect valuable information. The data model created from the existing databases and the information you obtain from the actual data by profiling it are of even greater value than existing documentation. Even lightweight profiling on the attribute level can provide significant insight. By looking at an attribute and its type (e.g., numeric, alpha, alphanumeric), examining the range of values across all records, and counting how many different values exist, one can grasp the meaning of this attribute. By looking at the number of different values and the number of null values, one can find candidates for a key. With this kind of profiling, even undocumented databases can be understood.

From the information gathered, a conceptual data model can be built within days. Check the accuracy of the model with the product owner of the analyzed IT system. If you find constraints within the model, write them down. Keep a link to the existing system and its data model so that before releasing every new version of the IT system, a check of the actual model against the conceptual model can be implemented and run. Ensuring that the conceptual data model stays up to date is a very good way to make sure future development will not go astray.

Deliver Insight and Collect More Knowledge Through Collaboration

Once prepared with the insight from profiling, all participants will find the discussion about the conceptual data model insightful; this will help ensure buy-in. Putting the conceptual data model on a collaborative platform will help as well. On this platform, everybody can add comments, ask questions, and be automatically notified about changes. Don't use this as a replacement for meetings; rather, employ this to prepare for meetings by answering a few questions: Were people active? What questions did they ask? On which parts did they comment?

In this way, work happens on all sides, even between meetings. Such collaboration should help generate excitement for your digital business initiative. Through the input received via the collaboration platform and the creation of the conceptual data model, you've already done most of the conceptual and design work on your data warehouse. The data warehouse then creates the necessary insight for evolving your digital business in an ever-changing world.

Add the Customer View

The conceptual data model is static. It doesn't tell anything about the processes that collect the data or in which order the transactions are carried out. It is only about the results of transactions and only about your own current data.

All the dependent business objects are customer touchpoints with your company. These touchpoints provide a picture of a customer's current documented interaction with the company, a very high-level view of how a customer experiences the company. A customer's interaction with the company resembles a process with each touchpoint acting as a task that results in data — your current dependent business objects. Participating business people are likely to think of possible immediate improvements to the customer experience now that they can follow through the data.

However, the picture is not yet complete. Only those touchpoints that have generated data are documented. Some touchpoints, such as when a potential customer first becomes aware of a company or its products, hardly result in any data, and no information is stored. To complete the customer picture, we must determine which interactions are not being documented and how to measure our performance on those interactions. What is not documented cannot be measured, and what cannot be measured cannot be improved. If a currently undocumented touchpoint is important, we need to determine whether we can obtain the data from external sources, and, if not, whether there is another way to collect it. With this knowledge, we can now add other (external) data to our conceptual data model. What is the customer's journey? What are the touchpoints during the phases of awareness, favorability, consideration, intent to purchase, and conversion?

Create a High-Level Business Model

A complete customer view provides a full picture of customer touchpoints, from first awareness until that person is sadly no longer a customer. The customer journey is not a process with a known beginning or endpoint. It is just single steps, some of which are dependent on prior steps. For example, a customer needs to complete the step "buy product" before the step "contact customer care" can happen. It must be clear where in the lifecycle of being/becoming a customer each step happens. And for each step, it must be evident whether there is company data or not; if data is there, the step directly links to the conceptual data model. If there is no data, a decision must be made about whether this data will be collected in the future (or maybe there is sufficient external data available in order to avoid additional development).

This process model of the customer journey will be put on the collaboration platform so that everybody can see and collaborate on improving the customer process and associated data collection. The model provides a living picture of the status quo, which can be modeled in Business Process Modeling Notation (BPMN) 2.0,[2] since it allows for complex processes where the tasks are neither mandatory nor executed in a certain order. BPMN 2.0 allows data objects as output.

Find Possible Improvements and Evolve

The conceptual data model, updated with the complete customer journey, makes it easy to find potential improvements in the business model and to align them with company strategy. The model can help determine whether a new development will bring the desired results. The initial Babylonian confusion is gone. Even though each participant still has a personal view on the business, he or she understands the differing views of other participants. Everyone can now communicate in a common language and work together to find the best possible solutions. Resistance and pushback should be low because people are heard. The company can function as a whole, utilizing its complete brainpower to find the fastest way to become a digital business. It can quickly achieve its first successes by aiming for the lowest hanging fruit; these first successes create the willpower to achieve and the budget to fund larger changes.

Integrate the Change

Change is likely to result in the addition of external data sources. As soon as this external data is on the company servers, it should be added to the conceptual data model as well as to the originating task/touchpoint. These changes will be visible on the collaboration platform. As the entire company sees the business model and the data behind it evolving, people will generate different ideas on how to proceed. Good, we need the best; test them, prove them, and apply them.

If the change affects existing IT systems, it should be visible in the customer journey as well as in the conceptual data model. The change will first appear as planned development; later, it will replace parts of the model as the new system goes into production.

When it comes to the Internet of Things, in particular, it is common for a product to generate information that varies depending on the version of the firmware. This should be documented in the conceptual data model, too, along with any limitations (e.g., limited usage of the data because of legal reasons). Restrictions are often among the first things forgotten.

Get a Living Model of Your Business

Now the customer journey and the conceptual data model are alive and evolving, rather than becoming outdated. In the past, a conceptual model often failed because changes were not integrated back into the model. To avoid this pitfall, the checks with the conceptual data model and the customer journey must be integrated into the release management of the software development process for the company's IT systems. This small effort produces a huge impact.

The two high-level models — the conceptual data model and the customer journey, which contain a lot of information — are very good requirements documents. The information can be used as metadata to the new or evolving IT systems, especially in development of BI systems, where this metadata has been used in the last few years to automate BI development. In my experience, this can decrease development time of BI systems by as much as 25%.

Conclusion

Digital business is all about data, so maintaining a conceptual data model provides a company with a vocabulary to address this data, thereby enabling individual employees to work with the data. Sharing the analysis work and the resulting conceptual model with the data warehouse development process saves time and effort. The more data is used, the more collaboration and input for change affect the data, and the more it becomes a living, evolving asset. Binding the conceptual data model to customer touchpoints creates insights into the business model. This business model is known to everyone and evolves with the business over time. There are two important factors to success: people and metadata.

With knowledge and a common vocabulary, people have two powerful tools to do their job better. Involving people and listening to them ensure their participation. Having a very clear picture of the current situation may sometimes be painful; however, accepting the status quo as the best possible solution so far — not imposing solutions — ensures collaboration on improving the current situation, a process that is now easier because of the new tools.

In the past, conceptual models often failed because of their tendency to become outdated. Using metadata for a vertical integration between the conceptual business models, the logical IT models, and the actual IT system solves this outdating problem. As a bonus, the metadata can be used to speed up the development process, especially in BI where data warehouse automation has been successfully implemented. The key factor for data warehouse automation is the metadata provided by conceptual models. Digital business is indeed all about data, and by creating a common vocabulary for the data of your organization and by visualizing the customer journey, you can finally talk about it and be understood.

Endnotes

[1]Greenfield solutions are similar to greenfield situations and greenfield projects. For more information, see: "Greenfield project." Wikipedia (https://en.wikipedia.org/wiki/Greenfield_project).

[2]"Business Process Model and Notation." Wikipedia (https://en.wikipedia.org/wiki/Business_Process_Model_and_Notation).

Michael Müller has been a BI consultant for nearly 20 years. He focuses on data modeling, data architecture, data vault modeling, and data warehouse automation. Currently, Mr. Müller is expanding into architectures and collaboration on data strategy/requirements for BI. He is on the board of directors of the German-language Data Vault User Group. He can be reached at michael@m2data.de.

How Metadata Can Drive the Move from Data Warehousing to Digital Business

by Pat O'Sullivan

Many of the principles that were developed for the data warehouse[1] are just as relevant today when considering the evolution to a digital business. Specifically, the focus on standardization and conformity to deliver a consistent view of information to many lines of business (LOBs) has parallels when considering the need to ensure a consistent experience for customers and partners when using an organization's different digital channels. Underpinning these principles is the implicit need for the active use and management of a coherent layer of metadata.

Using metadata to underpin various parts of the IT infrastructure is not new.[2] The capability to build out from a single canonical reference framework for different IT systems, including data warehouses and the applications that feed them, has been a goal for many organizations for decades or, at least, a goal for the enterprise architects within them. Unfortunately, in the past we saw the balkanization of many IT organizations into separate and often totally independent entities to deliver such systems as the data warehouse, the separate operational systems, and the services layer. Developers of these systems often only paid lip service to whatever enterprise-wide framework of metadata and designs was in vogue. So what is different now?

This article explores the latest thinking on how to evolve a common metadata system that truly begins to provide a key integration point between the data warehouse/data lake and the growing set of digital processes. It explores the components and characteristics of this new metadata layer and how it is essential to fueling the growth of artificial intelligence (AI) capabilities across the enterprise.

The Data Warehouse's Evolving Role in Digital Business

In discussing the broader digital business, it may help to delineate the various systems and applications that occupy the landscape of such a business. As shown in Figure 1,[3] a high-level classification of such systems could be:

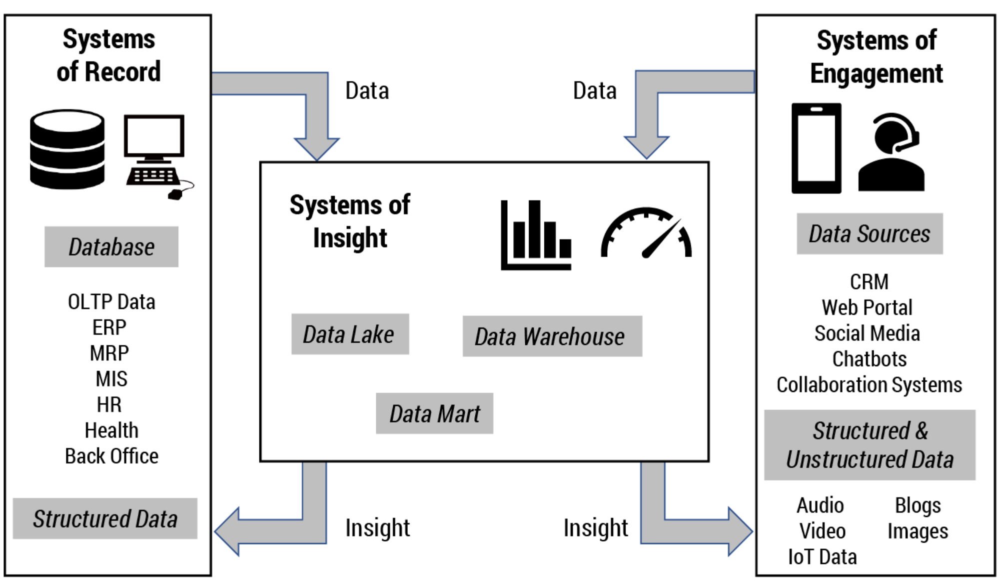


Figure 1 — The integration between systems of insight, systems of record, and systems of engagement. (Adapted from: Chen et al.)

- **Systems of insight** — data lake, data warehouse, and other associated analytics/business intelligence applications
- **Systems of record** —ERP and other back-office transactional systems
- **Systems of engagement** —Web portal, chatbots, and other applications used to engage with clients and business partners

The data warehouse was a response to the need to integrate and standardize key data and associated metrics from across a growing and divergent range of upstream systems that run the business — the systems of record. Such a one-stop shop of curated, managed, classified, and cleansed data provided the perfect basis for the creation of a multitude of analytics and reports to support many internal and external business needs, from various LOB reports to executive dashboards to the provision of data to various regulatory bodies.

A significant evolution in recent years has been the blurring of traditional frontiers between the data warehouse (or data lake) and the different systems running the business.

The delivery of these different business needs was felt to justify the often time-consuming and costly efforts to create and maintain data warehouses. In most cases, there was also a very clear delineation between the world of systems that ran the business — the systems of record and systems of engagement — and the world of the data warehouse and analytics — the systems of insight.

A significant evolution in recent years has been the blurring of traditional frontiers between the data warehouse (or data lake) and the different systems running the business. This is driven by the demand for embedded analytics in the digital business as well as the need for the data management systems to be more responsive in processing real-time or streaming data. One outcome was to dispense with the formal ETL layer that was the de-facto representation of the boundary between the data warehouse and these other systems.

Another change is the growing need for the delivery of automated capabilities across the enterprise, such as the expanding role of chatbots in the provision of a delivery channel; the birth of completely new business areas such as vehicle telematics in insurance; the availability of highly personalized experiences to customers based on their past purchasing history, expressed preferences, and current location; the increasing need for on-the-fly automatic analysis; and the need for classification of data as it enters the data warehouse or data lake.

Finally, one of the most significant common overall trends to emerge is the role of AI to revolutionize systems deployment across the technology landscape, driven by the need to deliver increased levels of automation and enhanced adaptability, as well as to address new use cases or business needs.[4] The growth of such automated and cognitive systems drives the need for more expressive and adaptive forms of metadata to enable and underpin such AI, which, in turn, raises questions about the traditional role of such metadata components as the data catalog, the business vocabulary, and the data model.

The Role of Metadata and Modeling Activities

In the past, explicit schemas or models assisted IT and IT-savvy users when understanding, navigating, using, and extending the systems for which they were responsible. This approach applied to the various systems that ran the business as well as to data warehouses and other analytical systems. With this predominance of predefined schemas, the data model took on a major role in enabling organizations to make sure that the specification of such database schemas reflected, as far as possible, the business and associated IT objectives that underpinned and funded the development of the data warehouse.

However, the advent of big data technologies with an emphasis on the ease and speed of ingestion of large amounts of data into a data lake — as opposed to the often-complex traditional ETL processes for loading into a data warehouse — meant far less focus on defining schemas or structures. Many data sets being loaded into the data lake either have no schema or bring their own implicit schemas. With many early data lakes, this sacrifice of structure to enable fast and easy ingestion was acceptable because the focus was either experimental or was initially intended to supply data to data scientists in support of their predominantly discovery-related activities. However, as the data lake becomes more mainstream and required as part of an overall data management infrastructure serving analytics with improved data provenance, the focus now shifts toward how to achieve an adequate level of governance of such data lakes. This is where the data catalog provides a central canonical reference point of business meaning to underpin any data governance activities of the data lake.

Physical Conformance vs. Catalog Conformance

One way of thinking about this is as a shift from traditional physical conformance to the achievement of conformance via the catalog (see Figure 2). Essentially, one of the traditional objectives of the central warehouse is to achieve conformance of the data by ensuring it is physically stored, subject to standard schema specifications; in other words, to achieve *physical conformance*. With the data lake, such conformance of data to a standard schema is neither desirable nor possible in any economic way; indeed, many big data/data lake exponents would call out as an anti-pattern any attempt to enforce schema standardization on data stored in Hadoop or other NoSQL formats. So, if the physical conformance of the data structures is not possible, how can an organization achieve any degree of governance of the data lake? One option is to consider achieving conformance via *catalog conformance*.

With catalog conformance, the objective is to ensure that the data within the data lake is mapped to an overarching catalog of interrelated business terms that effectively provide the business language used to describe the contents of the data lake. Thus, the contents of each data set within the data lake (or portion of it to be subject to data governance) are mapped to this catalog of business terms. Any users or applications accessing the data lake can use the catalog as a basis for determining the location, relationships, quality, and other characteristics of those data elements.

As organizations define and build this emerging business vocabulary for a very fluid and interrelated data lake, the question arises around the scope of such a business vocabulary — specifically, whether this catalog of terms is just concerned with the analytics area or whether it needs to have a wider scope, even to the extent of acting as the vocabulary for the whole enterprise.[5]

A critical aspect of this new approach to the provisioning of metadata is the existence of a layer of business metadata separate from but integrated with the metadata describing the physical details of the underlying structures. In the past, data warehouse metadata was often so focused on the data's physical aspects that it was of little use to people trying to build other non–data warehouse applications.[6] However, the evolution of data management metadata solutions now includes a semantic layer more focused on business language and business constraints.[7] This

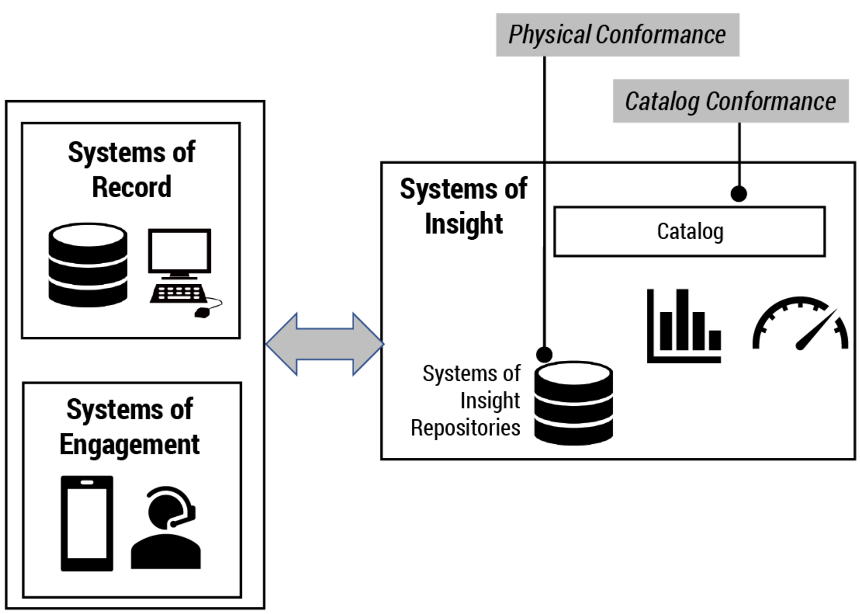


Figure 2 — Physical vs. catalog conformance in systems of insight.

means that the potential opportunity arises to use this metadata layer across the broader digital business.

In considering a broader scope for any data lake business vocabulary, it is often necessary to decide whether the scope is to just represent the business language of the data lake/systems of insight or whether it extends also to represent the core business language of other systems across the enterprise. Such a decision fundamentally impacts how this business vocabulary is designed, what personas and activities it is assumed to support, and what technology should be used to underpin the catalog.

For example, one approach is to build out a more expressive semantic layer of metadata with more extensive ontologies rather than limiting the metadata to the more traditional taxonomies or glossaries typically found in data management metadata repositories. Such ontologies have a growing track record of practical use in underpinning the Semantic Web,[8] as well as being viewed as key to the recent evolution of some cognitive systems. In this case, the focus shifts toward a metadata layer that is not just centered on the business meaning of the data elements and the type hierarchies needed for systems of insight, but is extended to cover the business rules, extended constraints, and relationships needed to reflect the additional needs of the systems of record and systems of engagement. Such an ontology may thus start to record the range of entities and intents needed to underpin the logic of chatbots, or the specific type hierarchies needed to support document discovery by customer agents.

In some cases, a pragmatic decision may be taken initially to limit the focus of any such business vocabulary to just the systems of insight. In other cases, the decision may be to define a cross-enterprise business vocabulary. In the latter case, a critical success factor of this broader use of metadata across the enterprise is the role of a chief data officer to own and champion the necessary cross-LOB cooperation, funding, and process of governance and change management.

Another alternative is a looser collaboration where organizations responsible for the different systems may use portions of this business vocabulary, but no significant effort is expended to enforce standardization across the enterprise. In fact, in many cases, this last approach becomes the default as the different individual developments progress in delivering point solutions across the systems of insight, record, and engagement.

The Future Potential Delivered by AI

Whether it is looking at the new function of metadata in the systems of insight or in supporting other areas of the emerging digital business, the increasing use of various AI capabilities is completely changing the role and usage of metadata. One common pattern is the shift from metadata being defined top-down to bottom-up. Specifically, there is an effort to move away from metadata being predominantly owned by information architects or enterprise architects to underpin mainly design-time activities with subsequent top-down deployments of metadata artifacts to the various runtime systems. Instead, there is a bottom-up demand for metadata to address specific areas, whether that is a vocabulary to govern a data lake; a type hierarchy used to assist natural language processing (NLP)–based extraction of key terms, obligations, and actions from unstructured data; or the need for entities and intents[9] to be used in building the logic flow of a chatbot. In many such cases, the demand for metadata is coming from specific point projects within departments, as the relationship between central IT and the business changes, and as such departmental efforts gain greater levels of autonomy.

Another common pattern is the ever-increasing presence of AI applications and their growing demand for high-quality metadata. Examples include the need for an effective structured-type hierarchy to support the processing of any NLP and the need for extensive and rich metadata of the data lake to support any AI-based discovery activities.

A major question is determining how to achieve a coherent strategy for the use of metadata across the enterprise to underpin such AI-based activities. There are benefits to defining an integrated framework of metadata that spans all aspects of the digital business. Building some of the aspects of a metadata framework can be expensive and may require specialized skills (e.g., the creation of an ontology needed to support many of the AI applications), so it may be more economical to do so centrally for many parts of the business rather than doing so piecemeal. Also, a centralized or at least coordinated approach means more reuse and commonality of such metadata artifacts across the business, resulting in fewer potential inconsistencies. In some cases, such an integrated approach to the use of metadata across the full breadth of the digital business may not just be beneficial from a business efficiency perspective but necessary from a regulatory compliance perspective. Increasingly, regulators are demanding a holistic approach. For example, the regulation on risk data aggregation from the Basel Committee on Banking Supervision[10] doesn't just demand the provision of the necessary reporting data but insists on the demonstration of the provenance of such data right back to its source in the systems of record. Another example is the European Union's General Data Protection Regulation (GDPR), which requires a holistic approach to the storage and protection of personal data across the systems of insight, systems of engagement, and systems of record.

Future Metadata Frameworks Offer Enhanced Digital Business Solutions

The increasing importance of a coherent organizational strategy to maximize the exploitation of data for growing a digital business is clear. There is a need to locate, consolidate, classify, and access the necessary data — often data of many different formats stored in different areas of the enterprise as well as external data — to drive many different aspects of a successful digital business. Examples include the increased use of AI-driven health monitors in hospitals to enable the early detection of combinations of potentially harmful symptoms; the ability of banks to offer more advanced wealth management applications that include access to far deeper research or provide a service to a wider range of potential customers; and the improvement in insurance of catastrophe risk via the use of machine learning models to automatically assess the severity of damages and predict repair costs of property based on historical data.

To achieve these types of capabilities in an efficient and scalable manner, the digital business needs to be able to identify, extract, transform, contextualize, and distribute the necessary data. The data warehouse or data lake is a critical part of this data ecosystem, and many different aspects of the digital business can exploit this resource in a consistent way by means of a shared metadata framework. Such a framework must be accessible and meaningful to the full range of business and technical users. It must be able to react to the inevitable changes that will occur in the business circumstances of the enterprise, such as changes in behavior of key customer demographics, the entrance of new competitors, and the impact of regulatory changes. This framework also must adapt to likely technological advances — not only the continued growth in the application of cognitive technologies but the potential advent of other applications in areas such as blockchain or the increased use of Linked Open Data[11] to grow the knowledge base of the digital business.

While the need for expediency in achieving short-term goals is understandable and at times necessary, the adoption of such pragmatism is only scalable in the long term when allied with an evolving cross-enterprise metadata layer. This is especially true when one considers that change will not happen in just a single area but is likely to come from a range of disruptive events across a range of technical and business dimensions. Such a level of change places a further premium on the ability of the digital business to react in a holistic manner.

> *The increasing importance of a coherent organizational strategy to maximize the exploitation of data for growing a digital business is clear.*

It may be beneficial to think of this metadata layer not as one single collection of terms, definitions, constraints, and rules, but more as a web of different but interrelated networks of knowledge:[12] a network of *business knowledge* encapsulated in the systems of record, a network of the extensive *data knowledge* available in the systems of insight, a network of *user knowledge* from the systems of engagement, and perhaps a network of *canonical knowledge* to provide the central spine of metadata against which all other networks can be mapped and aligned (see Figure 3). This federation of different networks enables the growth of the necessary interoperability of knowledge across the digital business, including the data warehouse, but also provides the necessary degrees of freedom to the individual areas.

As the digital business continues to expand, such an integrated approach to the management of metadata across all its components is a logical extension to the traditional integrated approach to data first specified in the data warehouse 30 years ago. The benefits of the traditional data warehouse of conformance, consolidation, and consistency can now be extended with a more adaptable, scalable, tightly integrated, metadata-driven ecosystem to ensure that the digital business continues to receive the necessary fuel of data.

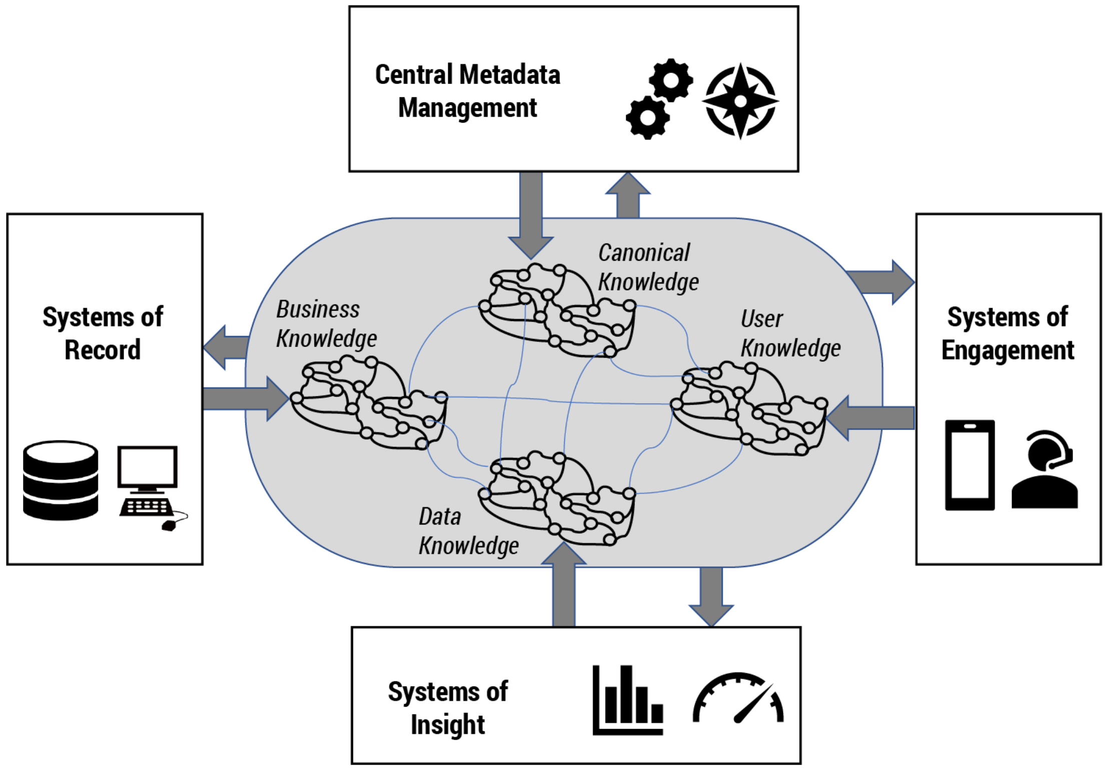


Figure 3 — The possible integrated network of knowledge across the digital business.

Endnotes

[1]Inmon, William H. *Building the Data Warehouse*. 4th edition. Wiley, 2005.

[2]Devlin, Barry. *Data Warehouse: From Architecture to Implementation*. Addison-Wesley, 1997.

[3]For a more detailed explanation of this classification of the different systems, see: Chen, Whei-Jen, et al. "Systems of Insight for Digital Transformation: Using IBM Operational Decision Manager Advanced and Predictive Analytics." *IBM Redbooks*, 2015 (http://www.redbooks.ibm.com/redbooks/pdfs/sg248293.pdf).

[4]Deshpande, Anand, and Manish Kumar. *Artificial Intelligence for Big Data*. Packt Publishing, 2018.

[5]Inmon, William, Bonnie O'Neil, and Lowell Fryman. *Business Metadata: Capturing Enterprise Knowledge*. Morgan Kaufmann, 2007.

[6]Gorelik, Alex. *The Enterprise Big Data Lake*. O'Reilly Media, forthcoming 2019.

[7]Mandy-Chessell. "Area 3 Models — Glossary and Semantics." GitHub, 2018 (https://github.com/odpi/egeria/blob/master/open-metadata-publication/website/open-metadata-types/Area-3-models.md).

[8]"Semantic Web: Ontologies — Vocabularies." W3C (https://www.w3.org/standards/semanticweb/ontology).

[9]Singh, Brijraj. "Chat Bots — Designing Intents and Entities for Your NLP Models." *Technology Evangelist*, Microsoft Developer Network, 29 January 2017 (https://blogs.msdn.microsoft.com/brijrajsingh/2017/01/29/chat-bots%E2%80%8A-%E2%80%8Adesigning-intents-and-entities-for-your-nlp-models/).

[10]"Principles for Effective Risk Data Aggregation and Risk Reporting." Basel Committee on Banking Supervision/Bank for International Settlements, January 2013 (https://www.bis.org/publ/bcbs239.pdf).

[11]"The Linked Open Data Cloud." Insight Centre for Data Analytics (https://lod-cloud.net/).

[12]Spahn, Michael, Joachim Kleb, Stephan Grimm, and Stefan Scheidl. "Supporting Business Intelligence by Providing Ontology-Based End-User Information Self-Service." *Proceedings of the First International Workshop on Ontology-Supported Business Intelligence*. ACM, October 2008 (https://dl.acm.org/citation.cfm?id=1452577&dl=ACM&coll=DL).

Pat O'Sullivan is a Senior Technical Staff Member for IBM Industry Models with 25 years' data warehousing (DW)/industry models experience. He is responsible for various aspects of models integration with such technologies as data lake, big data/Hadoop, DW appliances, ETL tools, and master data management. Recently, Mr. O'Sullivan has been leading an initiative to incorporate machine learning into processes to integrate the IBM Industry Models into the data lake infrastructure. Previously, he was involved for many years in banking models development and was one of the original team members that created IBM's Banking Data Warehouse – a suite of DW models that have been licensed by over 400 financial institutions worldwide. Mr. O'Sullivan holds a bachelor of science degree in computer applications from Cork Institute of Technology, Ireland. He can be reached at posulliv@ie.ibm.com or via Twitter @PatOSullivanIBM.

Cutter Consortium Membership
The Ultimate Access to the Experts

Cutter Consortium Membership opens up multiple avenues to interact with Cutter's experts to brainstorm and gain guidance to transform your organization and boost success.

Like everything business technology, one size does not fit all.

That's why we encourage you to choose the Membership that's right for your organization. Whether you choose Digital Transformation & Innovation, Enterprise-wide, Practice-specific, or CIO Membership you'll see a strong return.

Contact us at +1 781 648 8700 or sales@cutter.com to arrange a sample inquiry call with a Cutter expert and see for yourself how quickly your return on Membership can be realized.

What Do You Get from Cutter Membership?

- Get guidance in leveraging new strategies, emerging technologies, and business management practices to enable digital transformation and boost competitive advantage
- Learn how to mine data to create new products and services and improve customer experience
- Get input on how to reduce expenses through more cost-effective strategies
- Gain insights and get ideas on achieving sustainable innovation, successful change management, and prudent risk management
- Get coaching and insights on leadership and team-building practices that boost productivity
- Discover vendor-agnostic advice to ensure unbiased purchasing decisions

Test Drive Cutter Consortium Today

Have a question? A challenge you're trying to overcome? Have you reached a bottleneck and need some expert advice? Find out for yourself how valuable Cutter Membership can be — request a sample inquiry with a Cutter expert now.

Call +1 781 648 8700 or email sales@cutter.com.

Cutter Consortium
Access to the Experts

37 Broadway, Suite 1
Arlington, MA 02474 USA
Tel: +1 781 648 8700
Fax: +1 781 648 8707
Web: www.cutter.com
Email: sales@cutter.com

"I have personally been able to leverage Cutter's services since 1999. Among the attributes that differentiate Cutter from other firms, two remain at the top of my list: **Thought Leadership and Real Value**. Thought Leadership is driven by Cutter's experts. Real Value, driven by 'Access to the experts,' pushes us beyond just understanding the options. We can develop relationships with the experts, and tailor the options they present so that they can be quickly and practically executed within our organization, enabling our Business Technology team to continually improve, engage, and contribute to business growth."

— Doug Mikaelian
VP Business Technology,
Dairy Farmers of America

"The information technology revolution continues to change how businesses work and how people work in and with them. Today's demands on software to provide useful, usable, secure, and scalable services stress all aspects of systems development. These include architecture, technology selection, product development, and lifecycle management. Cutter provides comprehensive coverage for these crucial topics — through research reports and customized virtual training delivered globally — helping us to understand needs and solutions that span traditional silos."

— Eric Schoen
Director of Engineering, i2k Connect LLC
Former Chief Software Architect,
Schlumberger

"Thanks to Cutter´s experts, in a short time we have been able to:

- Improve our capabilities, using the contract scorecard for outsourcing
- Gain key insights and practical strategies for responding to challenges we face in building trust and partnerships within the organization and outside our institutional borders.
- Improve our methods for working with people and anticipating the challenges of cultural transformation within our organization.
- Have immediate access to key 'experts' at critical times, who spend quality time with our teams, provide state-of-the-art orientation, and help us implement projects.
- Implement agile, effective, and articulated innovation around a common vision for achieving quality education in Mexico."

— Arturo Cervantes
Director General of Information Systems
for Evaluation Results, National Institute for
Educational Evaluation, Mexico

CUTTER Business Technology Journal

About Cutter Consortium

Cutter Consortium is a unique, global business technology advisory firm dedicated to helping organizations leverage emerging technologies and the latest business management thinking to achieve competitive advantage and mission success. Through its research, training, executive education, and consulting, Cutter Consortium enables digital transformation.

Cutter Consortium helps clients address the spectrum of challenges technology change brings — from disruption of business models and the sustainable innovation, change management, and leadership a new order demands, to the creation, implementation, and optimization of software and systems that power newly holistic enterprise and business unit strategies.

Cutter Consortium pushes the thinking in the field by fostering debate and collaboration among its global community of thought leaders. Coupled with its famously objective "no ties to vendors" policy, Cutter Consortium's *Access to the Experts* approach delivers cutting-edge, objective information and innovative solutions to its clients worldwide.

For more information, visit www.cutter.com or call us at +1 781 648 8700.