

CUTTER

**Business
Technology
Journal**

Management, Innovation, Transformation

Data Architecture Is Really About People

Martijn ten Napel
Guest Editor

Personal and Organizational Context for Digital Business Decisions

by Barry Devlin p. 6

A Logical Information Architecture That Works for All Your Vs

by Adrian Jones p.11

Closing the Business-IT Gap with a Model-Driven Architecture

by Christian Kaul and Lars Rönnbäck p. 16

Practical Enterprise Data Architecture for the Digital Era

by Sagar Gole and Vidyasagar Uddagiri p.21

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. *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, artificial intelligence/machine learning, Internet of Things, blockchain, 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 monthly by Cutter Consortium, an Arthur D. Little Technology company, 37 Broadway, Suite 1, Arlington, MA 02474-5552, USA • Tel: +1 781 648 8700 • Fax: +1 781 648 8707 • Email: [cibtjeditorial@cutter.com](mailto:cbtjeditorial@cutter.com) • Website: www.cutter.com • Twitter: [@cuttertweets](https://twitter.com/cuttertweets) • Facebook: Cutter Consortium. ISSN: 2475-3718 (print); 2475-3742 (online).

©2019 by Cutter Consortium.
All rights reserved. *Cutter Business Technology Journal®* is a trademark of Cutter Consortium. No material in this publication may be reproduced, eaten, or distributed without written permission from the publisher. Unauthorized reproduction in any form, including photocopying, downloading electronic copies, posting on the Internet, image scanning, and faxing is against the law. Reprints make an excellent training tool. For information about reprints and/or back issues of Cutter Consortium publications, call +1 781 648 8700 or email service@cutter.com.

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

NOT FOR DISTRIBUTION
For authorized use, contact
Cutter Consortium +1 781 648 8700
or service@cutter.com.

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



by Martijn ten Napel

Opening Statement

International Organization for Standardization (ISO) standard 42010 defines architecture as “fundamental concepts or properties of a system in its environment embodied in its elements, relationships, and in the principles of its design and evolution.”¹ The many forms of architecture that are present today — whether digital, information, enterprise, business, or data — focus on a specific aspect of these fundamental concepts. Most architectures are design-oriented and, if you are lucky, they will guide you in evolving the IT landscape to keep up with the changes in the business IT supports.

But what exactly is a data architecture? I have asked myself this question many times over the past 20 years. What I observe is that data architects are busy designing systems that replace existing systems. The users perceive all kinds of problems with the information they get or have to record. They ask the data architect to rectify the problem, and the architect ends up replacing the existing solution with something new, in accordance with the latest fashion — or should I say fad — in data land. The end result is that the users’ problems are not solved, probably because the solution does not address the main cause of the problems in the first place.

Why is it so hard to design adequate data solutions and have the solutions evolve instead of replacing them over and over again? Over the years, I have started to see patterns across the different projects and businesses on and for which I have worked. One pattern is reliance on technology to solve data use problems. You could either see this as techno-optimism or as avoiding facing the real problems. The root cause of these problems is found in a lack of knowledge of how humans use information delivered by IT systems in their decision-making processes, as Cutter Consortium Senior Consultant Barry Devlin addresses in his article in this issue of *Cutter Business Technology Journal (CBTJ)*.

The technology industry, which has been promising technology solutions for organizational problems, has jumped into this gap. One of the rationales for the focus

on technology is actual technological progress. We have new data sources available, and the industry has grown tremendously with techniques to produce, aggregate, and analyze different types of data on a massive scale — what was beyond our imagination 20 years ago. Despite these capabilities, we haven’t progressed much in understanding how we humans deal with information, both in supporting business processes at an operational level and in using that data to create massive automated aggregation and analytical flows to guide participants in the business process.

Today, we use data to digitize more and more business processes, and with digitization we produce even more data that we try to leverage to our advantage. As Adrian Jones points out in his article, this increase in volume, veracity, and velocity leads us to vulnerability. Yet current data architecture practice fails to guide us in where we run risks and where we need to step up and organize ourselves to mitigate vulnerability. The disconnect between what a data architecture is in most organizations (a technology-focused roadmap) and what an organization needs from a data architecture (guidance on how to create a valuable and sustainable data landscape) has puzzled me for a long time.

Ever since the famous *The Economist* cover page a few years ago with the title “The World’s Most Valuable Resource” — depicting drilling rigs with the logos of Google, Facebook, and Amazon — people have misinterpreted that statement as saying that data is a resource, like oil.² But data is not a resource. Data is the breadcrumb trail of human activities. Like a true breadcrumb trail, it indicates the activity but never fully describes it. Data doesn’t fall out of the sky like manna from heaven; it cannot be mined like cobalt, either. Data is a residue of activity, and when you think about data in this way, you can envision a data management practice where you start to fully focus on the activities that bring you value and, therefore, the data you need to collect. All other data can be ignored; it costs you money to acquire and store, but it won’t bring you added value.

Data architecture should help you decide which activities are worth pursuing. The real challenge is that this is a dynamic system of loosely coupled activities. The volatility of changes in such a system varies in the course of time. Thus, a data architecture should be your guide on how to navigate these changes, which demands a lot from the interaction between data architects and the users of information. How should a data architecture support you?

For a data architecture to be an effective guide to how to manage data, it must be aware of the way humans use data. Even if we automate data aggregation, we always have to do so with respect for the use and for the fallibility of data and to guard against our own lack of understanding. The effect of relying on data in decision making is not without consequences. Virginia Eubanks has studied the use of data in social domains and has written a clear warning of how we fail people by deploying data solutions and algorithms.³ Though the examples she gives are related to injustices introduced into social support systems through relying on data and algorithms, the underlying mechanism in our business decisions is the same. We trust data-based decisions more than human knowledge-based decisions, even when we know that the data is flawed. I have yet to encounter a data architecture that accounts for the consequences of the use of data or has stern requirements regarding the consequences of data use. If mentioned at all, consequences are often viewed as a risk to be mitigated and not regarded as a fundamental property of a data architecture.

Creating an all-encompassing data architecture is not easy. A single architect is unable to understand all possible aspects and consequences. If you accept this and you are aware that data is the result of human activity — reflecting all the quirks we humans have in our behaviors — the answer is to create a framework where the value of using information is continuously evaluated against requirements derived from intended use. This evaluation leads to a continuous adjustment

of both the way you organize the use of data as well as the solutions that support that use. The technology roadmap is the result, with new technological capabilities an input to evaluate whether you should add better or new solutions to the landscape.

I call this framework the *connected architecture* (see Figure 1). The word “connected” emphasizes the necessary connection between different people, with different skills, who need to collaborate in acquiring, sorting, storing, modeling, processing, analyzing, interpreting, drawing conclusions, and taking action on the conclusions. You need to organize these connections; it won’t happen by itself. It is a continuous fight against data entropy, the scattering of copies of unmaintained data that result from local, one-time analysis and use. Data entropy is one of the plagues that visit many organizations in managing their data collections and in deriving value from them.

If architecture is indeed, as ISO standard 42010 defines it, the “fundamental concepts or properties of a system in its environment embodied in its elements, relationships, and in the principles of its design and evolution,” then connected architecture puts the environment in which data is recorded and used at the front of the stage — an environment consisting of people creating and using data.

In This Issue

Devlin sets up the theme of this issue of *CBTJ* by first taking us on a journey to help us understand how context plays a big role in using data. Known for creating the first data warehouse architecture, he proposed a new standard for data architecture for today’s world in 2013.⁴ Devlin puts context-setting information at the heart of all data architectures, and for good reason. In the drive to digitize more business processes, the intricacies of how all stakeholders interact with data have been underexposed. Though it is understandable that getting a grip on technology and reorganizing your business is hard enough, it is precisely this interaction that will determine your success. If you turn your perspective around, as he argues, your data architecture will be of more value.

So how do you put context-setting information at the center? In his article, Jones illustrates how to do this by putting Devlin’s architecture into practice. He stresses the importance of context-setting information by pointing out the increased vulnerability to which we are exposing ourselves. We produce and use more and



Upcoming Topics

Technology Trends and Predictions
Cutter Consortium

Digital Shift
Volker Pfirsching

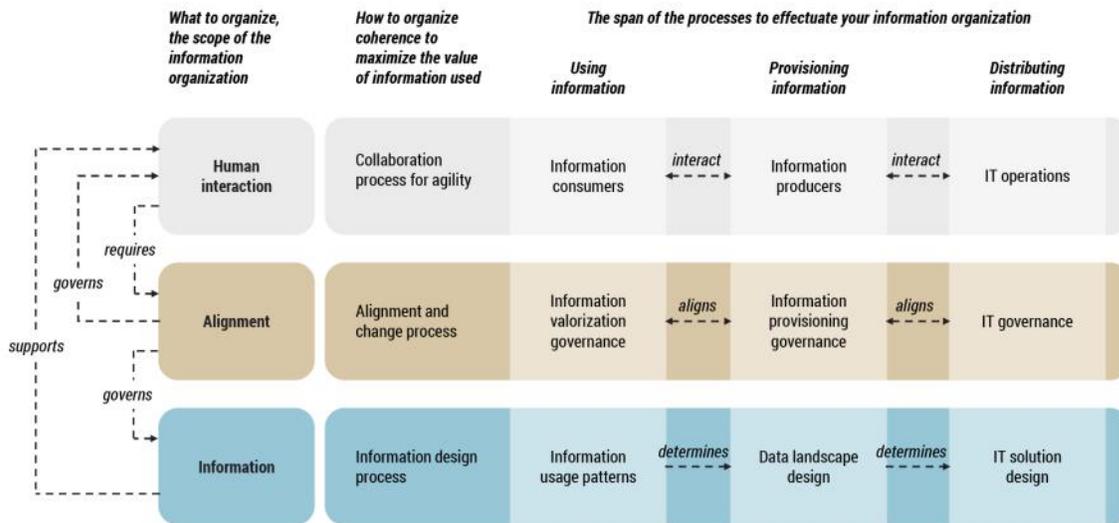


Figure 1 – The connected architecture framework.

more data. In the back of our minds, we know that data governance is of growing importance, but we don't act in the right way on this knowledge. The problem with data governance is that it is never part of a data architecture but rather addressed as a separate process. If our data architectures are not aware of the vulnerability being introduced, accidents are just waiting to happen. Jones hands you the recipe for avoiding these accidents.

You could, however, go even further. In the corners of the Internet, a debate about the need for a "data-centric" paradigm has been ongoing among data professionals. Every day, working data professionals face the risks and hurdles of data spreading around and being duplicated in different states of maintenance within an organization. Consequently, most data architects are fully occupied fighting the entropy of data and are desperately trying to make users understand that they are dancing on a rumbling volcano. In the next article, Christian Kaul and Lars Rönnbäck explore what it means to adopt a data-centric paradigm. It certainly isn't enough to have a data-centric data architecture; the implications are much more fundamental. The ultimate consequence is that you need to create a *model-driven* organization. By doing so, data architecture determines the shape of the organization, not the other way around. It's a thought-provoking article, and best appreciated when keeping the context of the two previous articles in mind.

So do you now feel invigorated to rethink your data architecture? If so, you need to prepare yourself for doing what is within reach of the capabilities of your organization now and what will result in a more people-oriented data architecture going forward. In

our final article, Sagar Gole and Vidyasagar Uddagiri help you understand which fundamental concepts — specifically, the six elements of an enterprise-wide data architecture — you should address today in order to "overcome challenges and leverage the opportunities and benefits of digital transformation." They describe the "secret sauce" that prepares your organization to thrive during a digital transformation journey.

I hope that this issue of *CBTJ* will help you understand that a data architecture should be much more than merely a technology roadmap. To be of any value to people in an organization, the architecture should guide the people in an organization to an understanding of how to organize for ever-changing information requirements.

References

- ¹ISO/IEC/IEEE 42010:2011: *Systems and Software Engineering – Architecture Description*. International Organization for Standardization (ISO), 2017.
- ²"The World's Most Valuable Resource Is No Longer Oil, But Data." *The Economist*, 6 May 2017.
- ³Eubanks, Virginia. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press, 2018.
- ⁴Devlin, Barry. *Business unIntelligence*. Technics Publications, LLC, 2013.

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.



Personal and Organizational Context for Digital Business Decisions

by Barry Devlin

As digital transformation — using data as the foundation — attempts to pervasively augment and automate decision making throughout business, understanding how people in organizations construct and use context to make good decisions is vital when it comes to defining a data architecture to implement a successful digital business. Yet the deeper we dig into “good” decision making, the closer we approach a new dilemma: *what might success in a digital business really mean?*

Before artificial intelligence (AI) conquered the world (or so it is claimed), analytics was going to unlock the potential of endless business insights from big data. Before analytics, business intelligence (BI) was going to replace the drudgery of decision making based on hand-worked data. Before BI, data warehouses were going to provide consistent access to all the data a decision maker could ever need. Before data warehousing, decision support systems (DSSs) and executive information systems (EISs) were going to revolutionize business management. Before DSSs and EISs,...

In truth, we know little about why some decision makers excel and how some organizations create viable decision-making processes.

Do you see where this is going? One technology after another has been touted over the past 30 years or more as the silver bullet that would empower and energize decision making in business. Each technical advance has offered, at best, small gains in productivity or ease of use for specific aspects of decision making but, overall, progress has been disheartening. Bad decision making by ill-informed people in dysfunctional organizations abounds. In the midst of all this technological advancement, we can still only find rare examples of excellent decision making.

Business schools and management consultants have offered theories as to why this might be so, with limited success. In truth, we know little about why some decision makers excel (spreadsheets aside) and how some organizations create viable decision-making processes. Nevertheless, we are embarking on a digital transformation journey that will pervasively automate decisions or augment decision makers via AI. As a result, the question put forth earlier takes on added urgency. If we don't know how our best decision makers or successful businesses arrive at good decisions, how can we hope to delegate some or all of the process to machines in a digital business?

Data Architecture Is the Answer. Now, What's the Question?

Within most enterprises, thinking about the process behind successful decisions is often left to the IT department. At best, this is probably not the ideal placement in the organization. At worst, sadly, the effort is confined to choosing tools to solve point problems or specific business needs and then trying to ensure that the supporting information is correct and sufficient. This post hoc attempt at data integration is what passes in many organizations for data (or information) architecture.

Suffice to say, this method generally fails to deliver more than transitory successes. Even if the business is happy with its initial solution, that state seldom lasts. New and changing needs gradually accrete around the successful solution as IT struggles to keep up. Business people invent and reinvent “new wheels” (usually in Excel), declare victory, and export them around the organization where they steamroll IT's architected offerings. That is, until it is discovered that cell AF534 references a data source retired two years ago, but some clever analyst has been maintaining a shadow copy, which works only if....

Data architecture as currently practiced is beset with a range of such problems, many of which are described in this issue of *Cutter Business Technology Journal (CBTJ)* — as are some solutions to some of the problems. However, I contend that we need to return to basics in decision making and ask how decisions are actually made by people and in organizations. In particular, a foundational question for data architecture that needs an answer involves the relationship between information and decision making.

The most widely used model describing the process of getting from data to decisions dates to 1989.¹ Management consultant Russell L. Ackoff posited a hierarchy with data at the base and tiers of information, knowledge, and understanding, leading to wisdom at the apex. Understanding is omitted from most representations — authors add or subtract terms to suit their needs — and the framework is widely known by the acronym DIKW. Ackoff suggested (ironically based on gut feel, it seems) that “on average about forty percent of the human mind consists of data, thirty percent information, twenty percent knowledge, ten percent understanding, and virtually no wisdom.” Definitions of the terms and the relationships between them have been continuously disputed since, and wisdom — both in the context of DIKW and in general — seems to be missing in action.

Personal Context and the Modern Meaning Model

In response, I defined the modern meaning model (m^3) in my book *Business unIntelligence*.² Understanding the difference between data and information and introducing the concept of context-setting information (CSI) are the first steps in addressing these disputes. *Information* is the physical representation of human communication and thinking, either created directly or gathered through machines, varying in degree of structure, and, today, primarily digital in form. *Data* is a subset of information that has been structured specifically for computational needs and consists mainly of well-bounded numerical and textual elements. Finally, CSI (or, simplistically, metadata) is also a subset of information that describes the relationship between data and information, as well as the disparate contexts in which information is created, used, manipulated, and disposed of.

An important consequence of these definitions is that information is conceptually a precursor of data, rather

than information being derived from data, as Ackoff suggested. Furthermore, the importance of context cannot be overstated. Successful decision makers are almost always expert in contextualizing data to correctly use information from various sources — a first step in understanding why they make better decisions.

The lowest *physical* locus of the m^3 representation shown in Figure 1 encapsulates the traditional scope of data architecture, showing the relationship between data and information, how they are sourced, and the transformation pathways between the two. Although not shown, CSI also exists in this locus, bridging the boundary between *hard* and *soft information*. Although m^3 was conceived mainly to show the path taken by business people from information via meaning to decision making/action taking, it should also be clear by tracing the directions of the arrows throughout the three loci that meaning must also trickle down to information. This “reverse pathway” occurs during the design and development of applications and data stores and is a particular focus for data modelers and architects. It is, therefore, a critical part of data architecture and should be kept in mind when discussing the two upper layers.

Moving from the physical locus of information storage and manipulation to the realms of the human mind and

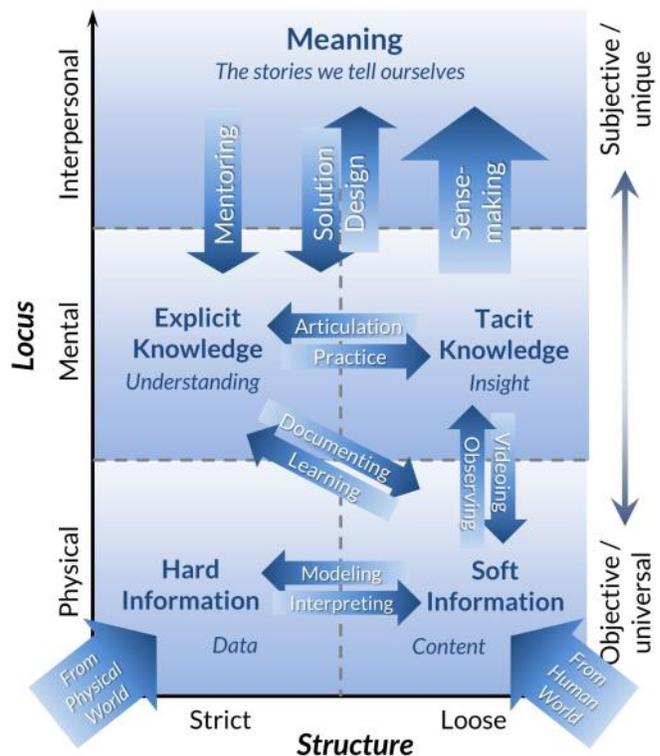


Figure 1 — The modern meaning model: m^3 .

social interactions is central to all decision making. Figure 1 shows these as the *mental* and *interpersonal* loci. The mental locus of m^3 , *knowledge*, shows that information must first be transmuted into tacit and/or explicit human knowledge before it can be acted upon. *Explicit knowledge* offers an understanding of the information available and an ability to document it directly in a usable form. *Tacit knowledge*, on the other hand, is more internal and difficult to document; the usual example of tacit knowledge is how to ride a bicycle. In decision making, tacit knowledge equates to the insight one has into the information that goes beyond the immediately available explicit physical information. The phrase “gut feel” is sometimes used to cover this area. The insight may come from prior digested information or internal mental heuristics informed or prompted by the incoming information. Good decision makers usually excel in moving between explicit and tacit knowledge, especially in relating prior and often unconscious information and context to the situation at hand.

Understanding and translating business needs into data-based systems requires communication skills, empathy, and trust between both parties.

The interpersonal locus, *meaning*, reflects the reality that we humans are, at heart, social animals and that business is a social enterprise. Understanding what information means — and, therefore, what actions are appropriate — always occurs in a social context. Meaning is, in the final reckoning, the stories we tell ourselves and others about the information we gather and the knowledge we hold. Decisions are seldom, if ever, fully rational. They are influenced by our emotional states, social conventions, and, especially, by our intentions and motivations, swayed by our own and others’ expectations. It is in this locus that leaders in decision making apply appropriate ethical and societal context to decision making, while non-leaders act with limited appreciation of the wider implications.

It is only when information has been internalized and integrated with knowledge and then transformed and expanded to meaning that valid and useful decisions can be taken, leading to successful actions and outcomes. And although it is stretching the basic meaning of the phrase “data architecture” far beyond its elastic

limit, for data architecture to succeed in a modern digital business, m^3 shows the broad scope — information, knowledge, and meaning — that it must encompass.

Returning to the reverse pathway mentioned earlier, it should now be clear that the initial interaction between business people (and their needs) and designers of data processing systems occurs in the interpersonal meaning locus. Understanding and translating business needs into data-based systems requires communication skills, empathy, and trust between both parties. In the absence of these characteristics — which is often the case — designers miss the subtle and often unspoken needs of the business people. The knowledge (in the mental locus) of the designers is thus faulty and overly influenced by designers’ prior assumptions or learnings. Documentation of even the explicit business needs is often deficient, and recording of the informal discussions, assumed background, and rejected solutions is usually nonexistent, leading to poor design of artifacts in the physical information locus.

Sadly, the situation is even more challenging than described here thus far. The reverse pathway is seldom a one-way street. No matter how good the initial design, business needs change and business people’s understanding of their needs changes with use of the system. Business people and designers must revisit the interpersonal meaning locus regularly to take account of such changes. In the absence of good documentation and participant continuity, these round trips serve mainly to build distrust between the parties. Contrary to popular opinion, Agile development — as currently practiced — struggles to address these fundamental impedance mismatches in meaning, knowledge, and information between business people and designers. While Agile solves the application-centric development problems of waterfall development, it is less easily applied to the challenges of data-centric development. Problems include the often enterprise-wide impact of changes in data definitions, the inherent research nature of BI and analytics projects, and the continuous and incessant nature of change in data needs in a digital business.

With these thoughts in mind, I have recently added a new pathway, *solution design*, to my longstanding m^3 diagram. This two-way path between the interpersonal and mental loci points to the need for ongoing collaboration both within and between business and IT in the conceptual and logical design phases for data-centric solutions. The continuous scope of such collaboration

and the extent of the parties involved — business people, analysts, data scientists, data engineers, and more — cannot be represented in m^3 directly, but rather emerge in the details of the *connected architecture* (referenced below). The goal is to strive to build explicit knowledge and create common tacit knowledge, which will lead to shared information meaning.

Digital Disruption of Decision Making

Digital business (or transformation) overly focuses attention on data with phrases like “data-driven decision making.” At the same time, it largely ignores the problems arising from applying vast swathes of often-suspect, biased, historical data to train AI algorithms. Furthermore, the twin emphasis on specific technological solutions and business needs distracts from the actual processes by which digital businesses could effectively shift from manual, human-centric decision making to an increasingly AI-automated or AI-augmented approach.

Applying m^3 thinking to digital transformation projects shows where the greatest benefits and dangers lie in automating or augmenting existing decision processes. We note that basic metadata and more extensive context-setting information in the physical locus is a *sine qua non* for even rudimentary confidence in the simplest of automated decisions. Without verified definitions and comprehensive provenance of all information used, automated (AI-driven) decision making will fall far short of the performance of today’s leaders who often use personal knowledge in this respect. Such personal knowledge would need to be harvested transparently as part of any digital transformation. BI and data-wrangling tool vendors today attempt to address these needs via collaborative functionality and text mining of informal human communications. However, this work remains at an early stage of development.

As we move to the mental and interpersonal levels of m^3 , the challenge increases significantly. Decision-making leaders operate largely on tacit knowledge and deeply held ethical convictions in these areas. Technology solutions are mostly ineffective in gathering such material. Rather, a focus on the actual decision-making processes, people, and organizational structures is required to determine the feasibility of “digitally transforming” these decisions and judging whether the underlying processes must be completely reinvented or if they should be left in human hands.

Organizational Context and the Connected Architecture

Given that the process of decision making at a personal level is still poorly understood, it might be unsurprising to discover that the process at an organizational level is even less well defined. This issue’s Guest Editor Martijn ten Napel describes the process as “riding the data beast” in an information-saturated environment riven with fragmentation, volatility, and context switching.³ Organization-level decision making must be agile enough to deliver boundless information in support of unlimited change in the market while simultaneously offering a solid and stable foundation of information governance to ensure ongoing coherence within and across enterprises. No easy task. However, ten Napel’s *connected architecture* (described in the Opening Statement of this issue of *CBTJ*) offers an initial framework within which we can explore and improve the organizational context for decision making.

As is the case with m^3 , the connected architecture also suffers from the fact that technology can address only a small part of its overall scope. It is people that matter most when we try to understand how organizations succeed (or fail) in decision making. In defining my own *ideal* conceptual architecture,⁴ I was faced with the challenge that there is no broadly based theory — rational choice theory being severely limited, in my opinion — of personal and social decision-making behavior that can be applied in the business realm. At the boundary of personal and organizational decision-making contexts, personal intention and motivation (both inherently unmeasurable) meet with social and organizational expectations (similarly vague) to create the frame within which meaning is defined across the many contexts in which the same or similar data is used across the business. This leads inevitably to consideration of how people collaborate first to set meaning and then to reach decisions and agree on actions.

As ten Napel says:

Human interaction is key.... 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.⁵

Technology has emerged, in particular from the social media world, that focuses on human interactions, relationships, and collaboration. A decade ago, such software (and hardware) seemed to promise advances in augmenting and automating the decision-making

and action-taking process. We have seen the resultant function included in many BI tools, but it is seldom of substance, focusing mainly on sharing questions and answers among team members. On the other hand, if stretched too far, it may begin to raise privacy concerns about employee surveillance and the use of personal information within the enterprise.

Some tools go so far as to use collaboration to “bridge the gap” between business and IT, a valid step forward. However, ten Napel suggests that we must view the process of creating meaning within context across three different organizational groups: (1) users of information who create value from it, (2) analysts who can provision and interpret information for users, and (3) IT who can automate data collection and prepare valid information from it.⁶ This is an insight that deserves further investigation and application in BI and data preparation tooling.

This and similar analyses of the organizational context of decision-making support suggest the complexity of this process is far greater than we have so far been able to describe. I conclude that concepts such as “self-service” and “citizen data scientists” — as proposed by some market analysts and software vendors — are, in reality, far too simplistic to form a firm foundation for future decision support strategies.

Final Thoughts

The concepts discussed here — m^3 and connected architecture — offer a new basis for considering how people and organizations can use and benefit from the information (and data) that is now harvested in such immense quantities from the physical and social worlds. They suggest that, at the very least, we may be placing too much reliance on the impact that such information has within real decision-making processes. Variations between businesses in personal intention and motivation as well as in organizational context for decision-making processes may, after all, be more important success factors than the amount and type of data gathered from external sources. Nonetheless, most organizations are likely to find some areas of possible improvement and others where information context is so lacking that the cost of fixing it outweighs the potential gains of transformation.

Such analysis will lead, I suspect, to the conclusion that a business cannot be totally digital in any practical sense.

Indeed, once we begin to take intention and motivation — at both personal and organizational levels — into consideration, ethical dilemmas about the meaning of success immediately arise. Analyzing the decision-making approaches and skills of accomplished decision makers within the business according to m^3 and the connected architecture may allow us to understand which decision processes are actually amenable to digital transformation. Can human skills be augmented in this process? Should human consideration be automated? Where lies digital right and wrong?

The question thus arises: *can we ever in good conscience conceive of a digital business without the beating hearts of human decision makers, however poorly we understand their decision-making processes?*

References

¹Ackoff, Russell L. “From Data to Wisdom.” *Journal of Applied Systems Analysis*, Vol. 15, 1989.

²Devlin, Barry. *Business unIntelligence*. Technics Publications, LLC, 2013.

³ten Napel, Martijn. “Taming the Data Beast: Inheriting Key Insights from Data Warehousing Practice.” *Cutter Business Technology Journal*, Vol. 31, No. 9, 2018.

⁴Devlin (see 2).

⁵ten Napel (see 3).

⁶ten Napel (see 3).

Barry Devlin is a Senior Consultant with Cutter Consortium’s Data Analytics & Digital Technologies practice. Dr. Devlin is a founder of the data warehousing (DW) industry, having published the first architectural paper on the topic in IBM Systems Journal in 1988. He is among the foremost authorities on business intelligence (BI) and DW. With today’s emphasis on big data, Internet of Things (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, he has guided BI and big data strategies and implementations in a wide range of business areas over the past 25 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 and, more recently, 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.



A Logical Information Architecture That Works for All Your Vs

by Adrian Jones

As data becomes an increasingly large factor in everyday living — overcoming organizations with a tsunami of new information — how can we cope with all this data alongside our legacy transactional data and make sense of it all? We need both solid transactional systems to run our organization and ACID (atomicity, consistency, isolation, durability)¹ database technology to ensure we are able to run our business transactions safely, consistently, and effectively. We also need to integrate and understand all the volatile data available to us in new formats, including unstructured data and data from social media feeds. No single system or approach is going to meet all our organizations' needs; instead, we need a hybrid approach with complementary components that can work together and an information strategy that supports this approach.

Where Are We Now?

A data integration strategy and information management within a sizeable organization can be incredibly difficult to define and even more difficult to implement. Almost all companies suffer from business and information silos, lack of governance, and fragmented legacy systems. Moreover, overcoming organizational, political, and sponsorship barriers is a huge task. As if these difficulties were not enough, the explosion of data around us and the need to make use of it to stay ahead of our competition further compounds the challenge.

So we need a strategy that allows us to reuse existing high-value components, a strategy that can integrate with new approaches and allow for flexibility, a growing necessity when it comes to future technological change and adapting to the “next big thing.” We must use our high-value data warehouse and ETL technology alongside newer platforms built to deal with the unstructured and big data world. Our relational databases and warehouses are not going away anytime soon; we just need to continue to maximize those investments.

For years now, the three defining properties of big data have been volume, variety, and velocity — or the three Vs. Several more unofficial Vs have been added over the years: veracity, value, and variability. These all make sense, of course, though these later Vs apply to all data, not just big data. The related problems in processing data existed before big data was a “thing,” and the age-old principle of garbage in, garbage out still applies.

Our data Vs, then, are *volume, variety, velocity, veracity, value, and variability*. And I might add another of my own here: *vulnerability*. How vulnerable is your data? What is your risk exposure from data misuse, misinterpretation, or leakage?

Others have added their own Vs to the list, particularly software and service vendors wishing to expand the list to facilitate sales. The actual number of Vs does not matter; we need to address all these data properties and apply good governance techniques to form a complete information ecosystem. We must support new development, outsourcing and cloud usage strategies, legal and compliance constraints, advanced analytics, self-service capabilities, and point-of-use integration of data. Thus, we need a logical architecture that can cope.

Our data Vs are volume, variety, velocity, veracity, value, and variability. And I might add another of my own here: vulnerability.

A Logical Architecture

Figure 1 details a design that utilizes Cutter Consortium Senior Consultant Barry Devlin's three-pillar approach² to managing data, while making the best use of available and appropriate technology platforms to process

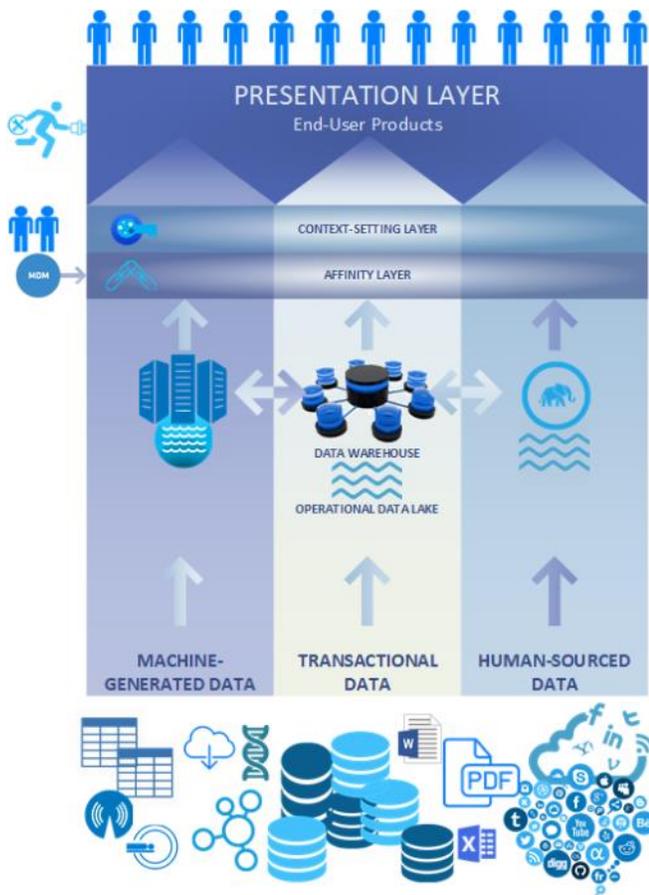


Figure 1 – Information architecture blueprint.

the three different types of data identified within an organization:

1. **Machine-generated data** — generated data, log data, device-captured data; high velocity, volume, and value; low veracity
2. **Transactional data** — traditional structured relational data from transactional business systems; high value; low (relatively) velocity, volume, veracity, and variability
3. **Human-sourced data** — unstructured, with high subjectivity; high velocity, volume, veracity, and variability; potential high value

This design also addresses the capabilities required to deliver end-user data services as well as providing integration possibilities throughout the chain from ingestion through to point of use. In general, Figure 1 depicts the flow of data from the source, at the bottom of the diagram, through to users at the top.

By using best-of-breed platforms to manage and deliver the data from its different source forms though to a

consumable data product, we can enable the integration of the differing data sources throughout the different processing cycles in the chain. It is necessary to integrate data from these different sources for multiple purposes, both in the structured and unstructured processing worlds, depending on the use case. The use case could mean moving data in or out of the relational world to answer key business questions using data warehouse techniques or data science and exploratory analytics. We have to enable this integration and make it as seamless as possible.

Importantly, the unstructured data lake and real-time data stream processing need to exist alongside the relational data warehouse and operational data lake, and they need to interact with each other seamlessly. And we must not forget that a data lake is nothing more than a storage area for collected data. We can think of Hadoop-based implementations as a massive shared drive. While Hadoop has many other uses for data processing, in its simplest form, the lake is, in effect, a storage area. In the traditional data warehouse world, the lake would generally have been your operational data store layer or data landing layer. A data lake approach in a modern organization could consist of both structured and unstructured storage. There are instances where having the structured, schema-on-write data available alongside rich metadata and lineage information is essential; this data can be mixed with data from the unstructured, schema-on-read type lake. You could think of this as having an operational data lake for all your internal, structured data sources, and a second lake containing all the unstructured, exploratory, data.

Data Diversity

The three-pillar data architecture (machine-generated, transactional, human-sourced) Devlin describes in *Business Unintelligence*³ is the basis for this logical design and accurately describes the data types we deal with on a daily basis. Overall, it is essential in a modern enterprise to process many diverse types of data. The traditional data warehouse is no longer able, on its own, to support the growing diversity of data types — both structured and unstructured — and the volume, variability, and velocity of modern data. A tiered approach maximizes the value of this data, with the appropriate approach and technology to extract value, while incorporating selectively chosen processed data into the data warehouse where required and where value is added, and vice versa from the data warehouse to other platforms.

Affinity Layer

The affinity layer provides the technical context of the data and related metadata, including underlying data-mapping capabilities to provide a layer of information to be utilized for the integration of source data from all pillars to produce target data sets. Data profiling and scoring are also part of this layer, and the outputs are referenced by the context-setting layer. This layer is essential to be able to work with integrated data from many sources, and particularly to enable integration at point of use, where the technical and business contexts are key.

Context-Setting Layer

The context-setting layer provides governance of the data, including business definitions, metadata, data lineage information and capabilities, policies, and data quality and profiling metrics. Regardless of the origin of the data, this layer ensures using the data in the correct context.

Data stewards and metadata stewards must manage the content of the context layer to build understanding of the data context and to enable the avoidance and understanding of data vulnerability and veracity issues.

The context-setting layer allows data consumers to answer questions such as:

- How do I find my data?
- Where did this data come from?
- What processing happened on this data?
- What policies govern this data?
- What is the quality of this data?
- What is the level of subjectivity of the data?

The context-setting layer is vital to understanding and to creating the big V, value. Without context, we cannot trust the data.

The Context-Setting and Affinity Layers and Why People Are Important

All tool vendors will tell you they have the tool of your dreams; that it will solve all your data integration and data management problems and will deliver data nirvana before you can pay the invoice. But the reality

is that these tools can only ever go so far. They cannot anticipate the irrational misuse of source operational systems and the human element in the capture and processing of data.

Much of the knowledge about the context of data that we create, update, and utilize in our organizations is tacit knowledge. Sometimes it exists only in the head of one individual who invented a workaround process, which was never documented, for something that the source system couldn't achieve. The tacit knowledge could be misuse of a system to achieve something entirely different from what the documented process describes. During my career, I have come across many examples when tacit knowledge derailed business intelligence (BI) and analytics projects.

Data stewards and metadata stewards must manage the content of the context layer to build understanding of the data context and to enable the avoidance and understanding of data vulnerability and veracity issues.

One recent example was where one particular field in an operational system, which was supposed to capture information about the regulatory status of a marketed product, was actually not used for that purpose at all. For all intents and purposes, the values in the field looked exactly as expected for the field's supposed purpose; however, the actual "meaning" of that data was completely different. In reality, this field was being used to capture a planning indicator, and the actual regulatory status data was stored in a completely unrelated field in the source system database. This misuse was due to the system having been shoe-horned into a preexisting business process. Relying on the technical mapping only, the output would have looked as expected but yielded completely the wrong result.

Without the intervention of people with knowledge of the processes and history, this misuse could have led to disastrous results in reporting to a regulatory body. Thus, there was a vital need for tacit knowledge in order to understand the context of that data and its actual usage, to document it, and to allow the technical team to create the right solution.

I have seen other recent examples in which data with several different meanings and contexts for usage was

stored in one single database field due to the business process requirements and the limitations of the transactional system. This storage misuse is something impossible to understand without input from the people with the tacit knowledge of that process and system.

The context-setting and affinity layers are the enablers of self-service and BYOT approaches.

Tools such as business glossaries and metadata management platforms fit into these layers to enable us to capture and document this tacit knowledge; surface it at the right time, to the right users; and ensure usage of the data in the right business and technical contexts. This becomes even more critical when using data in a data science environment, when its technical and business contexts can be transformed in a schema-on-read environment. The metadata is critical to ensuring the correct use of the data to develop hypotheses or solutions.

Context is everything. Garbage in, garbage out always applies. Data without context is just useless data. Data in context is information.

Bring Your Own Tool

Self-service analytics and reporting are key enablers in a modern enterprise, shifting the burden of development away from IT organizations and embedding the skills in the business. Technology enablement is essential to achieve success and to govern outcomes. The bring your own tool (BYOT) approach facilitates the use or “plug in” of any tool the end user wishes to work with to analyze or present data. This can include tools as diverse as Microsoft Excel, SAS, R, BI tools, and third-party mathematical or modeling algorithms. Enabling self-service creates value faster; however, self-service can only be managed effectively with good governance.

The context-setting and affinity layers are the enablers of self-service and BYOT approaches. These layers create the context and metadata required to achieve the level of traceability and governance required to work effectively with the data and actively manage proliferation of data copies. Although it is not possible

to physically control all actions of self-service users, particularly data copying and editing, centralized provision and access are key to successful adoption of policy.

Physical Implementation of the Logical Architecture

Many organizations already have multiple data warehouses and BI implementations and are also exploiting big data and unstructured data processing capabilities. Assessing the current corporate landscape for existing toolsets and data assets, which may already be leveraged to deliver components of the architecture, is an essential exercise. For instance, ETL tooling, a business glossary platform, a metadata management platform, master data, storage platforms, and so on, may already be available within the organization as a starting point. We must evaluate these and aim to reuse and maximize the value of existing IT assets rather than investing in completely new technology. We need to map those assets onto the logical architecture, assess the maturity of those areas, and identify where key opportunities and gaps exist. This approach will enable the building of a roadmap aimed at strategically building out the wider capabilities over time and moving from a base design toward a mature information architecture. In Figure 2, some of the current capabilities (at a high level) within a fictional organization have been mapped to the logical information architecture model.

It is also possible to identify the different source data repositories and systems for each pillar in order to identify high-value possibilities within current capabilities. The detail added at this logical mapping level will develop a view of the organization’s data and information processing capabilities and how they might be better leveraged.

The Future

The logical architecture is continuously evolving; this evolution must consider new developments and cultural shifts in thinking by organizations and individuals. Moreover, recent advances in, and commercialization of, artificial intelligence (AI) and machine learning technologies will undoubtedly influence the information architectures of the future.

AI, in particular, has great potential to remove much of the pain from the data preparation, ingestion, and integration stages, as learning algorithms can potentially handle techniques for these stages with a high degree of confidence in the coming years.

Similarly, the shift from old ways of data modeling and preparation will impact the data modeling approach. The impact on the logical information model and the need for human input at the context-setting and affinity layers will become increasingly important in providing good governance and the contextual confidence in the data to ensure its appropriate usage. We see a similar shift in the analytics world as the focus swings toward more probabilistic analytics outcomes and away from pure fact-based reporting.

All these new drivers require us to address our information management approach within our organizations and develop new approaches to exploit, govern, and manage these new developments as part of our overall information architecture.

Conclusion

By using this logical architecture as a base, we can understand how best to manage and process data in our organizations to derive value, and we can quickly identify gaps and opportunities.

Your organization can leverage the logical architecture as a guiding template to investigate and identify the current state of existing capability to deal with the three data pillars and with how to integrate them at key processing stages by:

- Identifying the data types in the organization
- Mapping current organizational platform and tool capabilities and looking for opportunities to leverage existing assets
- Identifying gaps that can be quickly and easily filled
- Roadmapping new and existing capabilities to the logical design to achieve an integrated ecosystem to support data usage

A final word: *there is no one correct answer.*

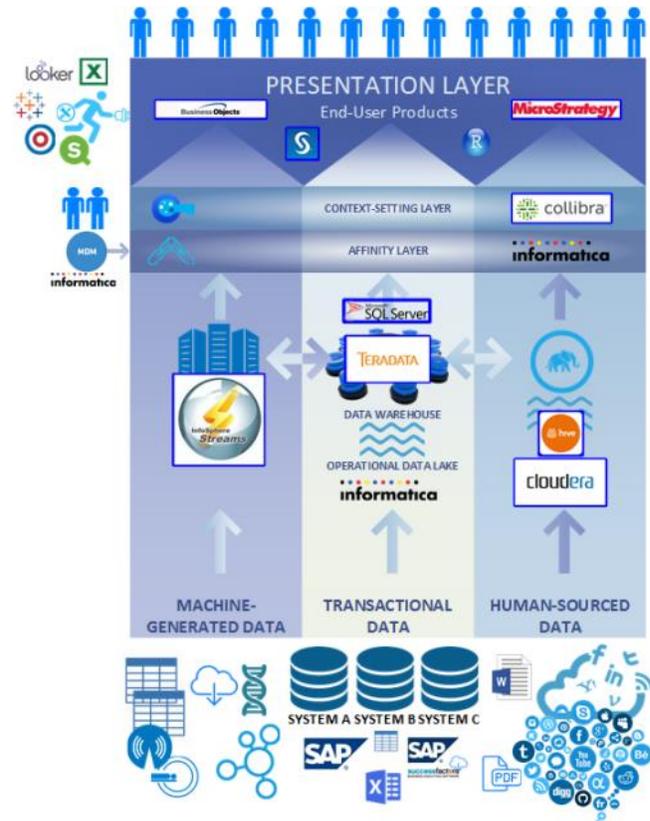


Figure 2 – Mapping your organization's capabilities to the logical architecture.

References

¹ACID (atomicity, consistency, isolation, durability) is a set of properties of database transactions. In the context of databases, a sequence of database operations that satisfies the ACID properties and, thus, can be perceived as a single logical operation on the data is called a "transaction." For example, a transfer of funds from one bank account to another, even involving multiple changes such as debiting one account and crediting another, is a single transaction.

²Devlin, Barry. *Business unIntelligence*. Technics Publications, LLC, 2013.

³Devlin (see 2).

Adrian Jones is Group Head of IT Strategy, Innovation, and Architecture at SYNLAB International. He is a highly experienced IT professional with almost 30 years' experience across software development, data and application integration, business intelligence and analytics, information management strategy, digital transformation, and enterprise architecture, primarily in the pharmaceutical and healthcare industries. He can be reached at jones_mailbox@yahoo.co.uk.



Closing the Business-IT Gap with a Model-Driven Architecture

by Christian Kaul and Lars Rönnbäck

You never change things by fighting the existing reality. To change something, build a new model that makes the existing model obsolete.

— R. Buckminster Fuller, 1982

Many organizations today struggle with a strong disconnect between their business model and their IT systems, data distributed over a large number of nonintegrated IT systems, manual interfaces between incompatible applications, or difficulties with European Union (EU) General Data Protection Regulation (GDPR) compliance. We can trace most, if not all, of these issues back to an abundance of unspecific, inflexible, and non-aligned data models underlying the applications these organizations use to conduct their business. Often, these data models have been developed with insufficient business involvement and in isolation from each other — or have been purchased from vendors with little concern for the actual needs of the organization.

In this article, we briefly describe some of these issues and their origins. Next, we introduce a practical team-oriented data modeling approach that allows you both to document how your business actually works now and to build a solid, easily extensible foundation for the processes and applications that will make it work the way you need it to in the future. Finally, we outline how this approach could be at the heart of a new kind of organization that eliminates the traditional business-IT divide. Instead of creating models that mirror the existing organization and its dysfunctions, we suggest first creating the model and then deriving the organizational structure from it.

Silos Upon Silos

Organizations usually have some understanding of the concepts that drive their business. But the entities used in their IT systems often don't align with these concepts. Furthermore, their employees are divided into departments and teams in a way that doesn't match the concepts either.

This disconnect leads to a vicious cycle of people silos creating data silos creating more people silos.¹ Leaders of different business units don't talk to each other and create data fiefdoms out of organizational fiefdoms. They fail to recognize that most of the undertakings they see as IT projects actually are business transformation programs. It's not enough to throw money and some requirements over the business-IT fence and then wait for the magic to happen. All stakeholders, business and IT alike, must be aligned to successfully complete the transformation.

Silos enforce the idea that data behaves much like a traditional manufacturing resource that needs to be physically present at the location of use. But data is different. At least in theory, it is a nonconsumable resource, instantaneously available everywhere and perpetually reusable. Unfortunately, organizations are often unable to leverage the unique characteristics of data. The process of continuous siloization found in many organizations, and in almost all organizations above a certain size, means that the data that employees need to do their work is distributed over a large number of nonintegrated IT systems within different intra-company jurisdictions.

Often, it is hard for people to find the data they think they need or to even know that the data is already there — in some system, somewhere — in another part of the organization. In other cases, multiple contradictory sources exist for a certain data element, without any clear indication as to which one holds the correct values. The result is that data about a single topic has to be stitched together manually from different systems with different data structures or with the help of large quantities of what is euphemistically called "business logic." With all these silos, compliance with the GDPR and similar data privacy regulations is close to impossible.

While it is difficult to quantify precisely the impact of the time people spend seeking, compiling, and harmonizing data in their daily work, it's obvious to most organizations that the cost, duration, and overall complexity of what they see as IT projects continue to rise.

While it's easy to blame the IT department's general incompetence for these issues, siloization is usually the root cause of the problem.

However, IT people are not totally innocent bystanders. Too often, they seem reluctant to work with the business side to determine the actual challenges the organization faces and instead keep to themselves, building system after system after system without really knowing whether these systems are what the users need. Sometimes, they even try to impose their preferred tool, or preferred vendor, regardless of the actual requirements, just because that's what they know and like.

Building all these siloed systems, organizations accumulate what Dave McComb calls "integration debt":

Integration debt occurs when we take on a new project that, by its existence, is likely to lead someone at some later point to incur additional work to integrate it with the rest of the enterprise.²

This kind of debt is what tends to make each IT project more complex and more expensive than the last. The number of nonaligned IT systems inside an organization keeps increasing, making it more and more difficult to stitch them all together without the whole thing falling apart.

Modeling Your Reality, Together

How can an organization escape this vicious cycle? After spending years on various kinds of data-related projects and having countless, and often depressingly similar, discussions with practitioners, we've come to the realization that the solution could be to put data modeling front and center. It just has to be done a bit differently than it has been done in the past. The new way we're about to describe won't have you bottlenecked by building complex enterprise data models in an ivory-tower fashion before implementing anything; far from it.

First, we have to realize that data models only become really useful when they also work as communication tools, documenting with sufficient detail how an organization works now and how it will work in the future. When a model captures both the current reality and the desired reality, it can serve as the core of a *model-driven architecture*, with all kinds of technical and nontechnical artifacts generated from it. In the process of creating such a model, the people in the organization develop what Eric Evans calls a "ubiquitous language,"³

a common vocabulary that makes sure that everyone understands what everyone else in the organization is talking about.

This common language is the first prerequisite for escaping the vicious cycle of siloization. With its help, an organization can overcome the Tower of Babel-like confusion caused by silo-specific dialects that use different words for the same thing or, even worse, the same word for different things. The second prerequisite is to stop seeing a data model as purely technical and specific to an application, such as something that describes the database of a particular system. Instead, think of a data model as a description of what actually happens in the organization, a model that is shared between applications.

How can you build such a model? Do not fixate on generic concepts early on, a trap people within IT often fall into, before you have a grasp of the individual cogs that make your organization tick. Start with interesting instances that exemplify your daily business, whatever it is. Don't talk about parties involved in a transaction at a location, talk about "Dominique buying a red T-shirt from Alberto at the Poughkeepsie store." Flesh out the instances with details such as age, marital status, or physical characteristics; think about the identifiers you might use to recognize the people involved. Think persona, not generic concept. Only after you have collected enough instances to cover the relevant parts of your business should you start thinking about how to classify them.

Finding instances and determining classifications isn't something enterprise architects or database developers should do on their own. On the contrary, it is a collaborative exercise involving stakeholders from both sides of the business-IT divide — a divide that we hope gets smaller in the process. You need to involve all relevant actors to make sure that the agreed-upon classifications really are common classifications, acceptable to everyone who works with whatever has been classified. Only then will you arrive at a data model that is specific to your organization's needs and supported by a broad consensus within the organization, not just a model that serves as the foundation for yet another data silo.

Keep in mind that these common classifications may change over time. Your organization and the way it conducts its business is evolving, and instances and classifications must evolve, too. By regularly working on instances and classifications together, the business and IT sides of your organization will make sure your ubiquitous language stays up to date and, in the

process, will gain a better understanding of each other and the state of the organization.

Generating the Next Levels

Many data initiatives put too much effort into creativity at the technical implementation level. Instead, the focus should be firmly on the business-minded discussion — first, of instances and, then, of classifications, as described above. All technical artifacts should be generated with minimal human involvement, freeing up time and energy for whatever is the organization's core business. Think data warehouse automation, but for all kinds of physical data stores instead of only for data warehousing.

Remember the three levels of data modeling: conceptual model, logical design, and physical implementation? The interesting instances and common classifications found in the collaborative exercise described above form a conceptual model from which a highly normalized logical design can be generated automatically that includes all the relevant:

- **Concepts** — the things that are important for the people involved
- **Connections** — the relationships between these things
- **Details** — the attributes of these things

This logical design holds everything together. It serves as the foundation for all the physical data stores that might be needed, making sure all of them speak the same ubiquitous language and are therefore interoperable with each other. Again, these physical data stores don't have to be painstakingly modeled by hand; they can be generated without much effort from the logical design using mathematical representations of standard modeling patterns, such as fifth normal form, data vault, or anchor.⁴

Deriving data stores from an organization-wide logical design contrasts with the traditional method of creating a dedicated model for each and every application. Most existing organizations are, in that respect, application-centric. McComb, however, defines the *data-centric* organization as "one in which most application functionality is implemented on a single, simple, update-able, shared, federate-able data model."⁵ Given that definition, data-centric actually might be a bit of a misnomer. More than merely being centered on data, the organization is driven by a single model. Just how

far we can take this somewhat radical thinking will be easier to understand if we first take a stroll down memory lane.

From Many Models to One

A quick IT history lesson will help us understand the current situation and the related paradigm shift about to happen. When computers were largely unconnected, it was necessary for data, interfaces, and logic to reside alongside each other. The purpose of the interfaces and the logic was to fetch, display, modify, and create data, usually with a human involved in the process. This meant that even within a single computer running one program, it made sense to separate concerns among these three elements. During the 1970s, such ideas were even formalized in the programming language Smalltalk-79 under the acronym MVC (Model-View-Controller), a design pattern that lives on in most modern programming languages.

With the widespread use of this pattern, it is somewhat perplexing that in a time when it is hard to say where one computer ends and another begins, or if they are local or in the cloud, the way we think about applications has changed very little. Applications are mostly the same monoliths as they were back in the 1970s, keeping the same single-computer architecture, but now with a virtual machine on top of any number of elastically assigned physical units. We have already gone from the computer as a physical asset to compute as a virtual resource, but this journey is now also beginning for data. Thanks to data being able to flow freely, applications can be built so that it's also difficult to say where one application ends and another begins, and applications may seamlessly run from the edge to the cloud.

Centralizing data is, of course, not an entirely new thought. Centralization has been conceptualized through master data management (MDM), in which the responsibility for a few important organizational concepts is taken away from the zoo of applications where it usually resides. While many MDM initiatives have begun over the last decade, most of them haven't added much value. We believe that from both a technological and a theoretical perspective, MDM may have been released into the wild prematurely. It requires proprietary languages to communicate, is slow to adapt to changes, and demands complete consensus on how to define the data for which it is responsible. Since initiatives centralizing just one concept have had little success, centralizing all of them might sound

impossible. But this is one of the problems where holistic thinking actually will turn the odds in your favor.

Because no organization can function with only a single application, good standards on how to exchange information electronically now exist, largely thanks to so-called service-oriented architectures. It is no longer necessary to build data models to last; they can be built for change and can be malleable with respect to differences in opinion about definitions.⁶ Finally, the meaning of “master” has been made operationalizable: a *master* is responsible for the identification process, via which it is possible to determine, given some pieces of information and through smart comparisons with stored data, whether something has been seen before or whether it is something new.

We are, therefore, now better equipped than ever before to go from application-centric to model-driven. There are no longer technical limitations preventing applications and the people working with them from speaking a common, ubiquitous language throughout an organization. When things change, the terminology used and the model with which people work can change with it. Probably the biggest challenge will be to get application vendors to release control of the data they have internalized. For many, this will involve major rewrites of their software, which doesn’t come easy or cheap. Will vendors that embrace a model-driven architecture gain enough traction to force everyone to follow suit? Only if the buyers — you and the organizations you work for — start to demand it and enforce it once in place.

From Model-Driven Architecture to Model-Driven Organizations

Even when an organization works with centrally identified pieces of information that fit into an organization-wide logical design, one important misalignment remains: the one between the general understanding of the concepts that drive the business model and the departments and teams into which the employees have been divided. To bridge this final divide, the organization has to make the leap from a model-driven data architecture to a *model-driven organization*; it has to derive its own structure from the logical design.

In a truly model-driven organization, the organizational chart is just another physical implementation of the common logical design. People will work together in small, cross-functional teams that are one-to-one with

the concepts that are important to the organization right now. If, for example, your important concepts are Customer, Employee, Product, and Sale, then you’ll have teams called Customer, Employee, Product, and Sale that are responsible for the respective concept, its details, and the physical data store(s) associated with it. The common business-IT divide will slowly become obsolete because, to fulfill all its responsibilities, each team will have to include both more business-minded and more technical-minded people. At the same time, the one-to-one relationship between concepts and models will prevent the reemergence of different understandings of the same concept in different parts of the organization.

Of course, none of these teams would or should be an island. There will be defined interfaces between the teams that are one-to-one with the connections from the logical design. Teams are jointly responsible for their common connections and the physical data store(s) associated with them, usually with one team in the lead. In our example, there will be a connection between Customer, Employee, and Sale, and another connection between the Sale and the Products that have been sold. In both cases, it makes sense that the Sale team takes the lead because Sale is the concept that ties all these other concepts together. These institutionalized connections will make sure that no team can isolate itself from the others and degrade into one of the people silos of old.

Everything put together, the *modus operandi* of such a model-driven organization would consist of these four steps repeated over and over again:

1. Regularly document representative examples of what happens in your organization (the interesting instances) to see whether your model still aligns with how you actually conduct (or want to conduct) your business.
2. If there are significant discrepancies between the model and the reality you want and you have to make structural changes to the organization, build consensus to alter the relevant classifications accordingly.
3. Generate the new logical design of your organization from the altered conceptual model (the current combination of interesting instances and common classifications).
4. Generate the new organizational chart and changes in the physical data stores from the altered logical design.

An organization that follows this process will have close alignment between the general understanding of the concepts that drive its business, the entities used in its IT systems, and the teams into which its employees are currently divided. Because physical data stores and organizational structure are generated from a common model, they can't stray too far from the business concepts. If discrepancies evolve over time, they are detected early, and the model, data stores, and organizational structure can be adapted easily.

While a traditional organization's overarching purpose may be well known to its workforce, that purpose is usually hard to translate into the rationales behind individuals' work and the particulars of daily operations. In a model-driven organization, everyone has a crystal-clear purpose; people's actions serve the purpose of fetching, modifying, and creating data about a specific concept. If their actions do not leave digital traces in the right places, they are not doing what they are supposed to be doing. In such an organization, it's clear where to find customer data because all customer data can be found in one place, shepherded by the Customer team. While there isn't one authority for everything, the team that is responsible for a concept is the one authority for everything related to that concept.

An organization will no longer have *the* single source of the truth, which it probably never had anyway, but it will have *a* single source of the truth for data about each concept that is relevant to the organization, and it will be obvious what this source is. This architecture not only helps the people working inside the organization to find the data they need and update it in the right and only place, it also helps people from the outside interacting with them. For example, regulatory authorities may require data from the organization, or customers may want to see, and sometimes change, the data the organization has about them (GDPR data subject requests may well make this type of request commonplace). In a model-driven organization, employees will be able to fulfill such requests faster, more easily, and more accurately.

Conclusion

Whatever you may think of its feasibility with respect to your organization, model-driven architecture is now something every architect must consider. It has been done in practice at companies such as Hitachi, Google, and Bitly, with very promising results, and can be done at other organizations. Ravi Ganesan, a proponent of

data-centricity, explains that "one powerful way to become a high-performing organization is to bend the power of technology to support organizational performance."⁷ Since every organization now uses technology to deal with data, just having lots of technology for handling data is no longer a differentiator. What is a differentiator is to what extent the organization leverages this data to advance its business. Having a model-driven architecture and becoming a model-driven organization are powerful approaches to successful data leverage. If your organization won't take the leap, it risks being overtaken by organizations that will.

References

- ¹Silverston, Len. "Zen and the Art of Data Maintenance: People Silos Cause Data Silos" *The Data Administration Newsletter*, 6 November 2019.
- ²McComb, Dave. "The Data-Centric Revolution: Integration Debt." *The Data Administration Newsletter*, 1 March 2017.
- ³Evans, Eric. "Domain-Driven Design: Tackling Complexity in the Heart of Software." Addison-Wesley Professional, 2003.
- ⁴Kaul, Christian. "From E-R to Data Vault: Transforming Models with Shape Functions." LinkedIn, 5 August 2018.
- ⁵McComb, Dave. "Data-Centric: Models and Architectures." *Data Architecture Summit*, 15 October 2019.
- ⁶Rönnbäck, Lars. "Modeling Conflicting, Uncertain, and Varying Information." ResearchGate, 2018.
- ⁷Ganesan, Ravi. "Utilizing Technology to Build a High-Performance Organization in a Value-Based Environment." The OPEN MINDS Performance Management Institute, 14 February 2019.

Christian Kaul is a data modeler, writer, and event organizer based in Munich, Germany, who focuses on designing, implementing, and maintaining data warehouses. He has several years' business intelligence experience in various industries, including healthcare, insurance, tourism, and telecommunications. Mr. Kaul's project roles have included data modeler, data warehouse developer, project manager, and support team lead. With a keen interest in new developments in the data modeling area, he organizes a local meetup group and chairs the Knowledge Gap data modeling and data architecture conference in Munich. He can be reached at christian.kaul@obaysch.net.

Lars Rönnbäck is a researcher in divergent information at Stockholm University, Sweden, and a consultant in information modeling. He has been working with some of the largest organizations in Sweden and their challenges as far back as Y2K. Over the last decade, Mr. Rönnbäck's research has focused on how to manage information that is uncertain, imprecise, conflicting, and varying over time. This research has resulted in new modeling techniques that provide a closer-to-reality record of events, most notably anchor and transitional modeling. He has also written a book about how young children learn to count. He can be reached at lars@uptochange.com.



Practical Enterprise Data Architecture for the Digital Era

by Sagar Gole and Vidyasagar Uddagiri

Enterprise data systems have been evolving continuously alongside computing in the five decades since the advent of the first IT business applications built on an IBM mainframe platform. This evolution has addressed changing business requirements while leveraging technology improvements in the market and continuously overcoming a plethora of business challenges. In particular, globalization has been a major challenge across the journey of continuous improvement. Yet the advent of the digital era is creating challenges far more difficult than globalization — or any other macroeconomic or societal phenomenon that humankind has ever faced.

Even so, the digital era has created many new possibilities and opportunities for enterprises, which involve developing insights from data using advanced analytics as well as data monetization. While some progressive enterprises have been leveraging digital technologies to enable disruptive change, many other enterprises are at different levels of maturity with respect to harnessing the real power of transformation that digital technologies enable.

Enterprises must manage their data systems to overcome challenges and leverage the opportunities and benefits of digital transformation. The “secret sauce” to achieving this lies in a practical approach that can help enterprises transition from their current state to the desired state in which they are ready for the digital era and beyond. One key element of enterprise data systems is enterprise data architecture, whose function is to oversee how business entities’ data is maintained, synthesized, and made available in a seamless manner across the value chain, cutting through internal boundaries by virtue of a practical governance mechanism.

An analysis of success stories from the digital era clearly shows that enterprises that have successfully transformed themselves have transformed their business behaviors. Four such critical behaviors include:

(1) driving mass personalization, (2) creating exponential value, (3) leveraging ecosystems, and (4) embracing risk.¹ Data and its architecture play a pivotal role in enabling these critical behaviors.

To drive *mass personalization*, the enterprise should have a solid data architecture foundation to collate data from various channels. The business itself may be operating in an omnichannel model. The ability to blend data from these channels and to blend that data with external data sources (e.g., social media feeds and external data provider feeds) is instrumental in driving mass personalization.

Enterprises must manage their data systems to overcome challenges and leverage the opportunities and benefits of digital transformation.

Now let’s consider an example of creating *exponential value*. A limited number of warehouses initially seemed to prevent a retailer from expanding its online operations across a wide geographical area. However, the retailer had stores across the area in question. So the business implemented a novel concept: the retailer started fulfilling online orders through its stores. This model, of course, had its own set of challenges, including that the retailer could sell out of an item in its brick-and-mortar stores, leaving it unable to fulfill online orders for the same item. The company addressed such challenges largely by leveraging data across channels and having the right data architecture, enabling the business to create exponential value.

Auto manufacturers provide a good example of *leveraging ecosystems*. They leverage digital platforms to partner with consumer electronics producers, entertainment firms, telecom companies, and other service providers

to support connected car offerings. To enable this behavior, the data architecture must be extensible enough to assimilate data from this variety of sources and then derive meaningful insights from it.

When it comes to *embracing risk*, enterprises need to analyze data from the enterprise, the ecosystem, and the market and use that analysis to assess risks and evaluate the likelihood of success after taking a given risk.

So What's the Problem?

Enterprises that are failing in the digital era are often unable to harness the abundance of data effectively. The reasons are primarily due to being overwhelmed by the complexity of data management and being unable to effectively handle data ecosystem challenges. These challenges arise due to lack of the following:

- Data standardization
- Data quality
- Data harmonization

Enterprises also face the challenge of dealing with heterogeneous data formats, diverse regulatory rules and industry standards, and the complexities around managing data governance.

What makes the task of choosing the right technology significantly more difficult is the plethora of technologies on the market.

Transforming People, Process, and Systems

The triad of people, process, and systems (technology) transforming cohesively are the ingredients for the secret sauce that can provide a practical way of managing the enterprise data architecture. Let's examine each of these influences on data architecture.

Technological Influence on Data Architecture Management

At a broad level, an enterprise data landscape comprises data integration, data persistence, data delivery,

and data governance. A decade ago, choosing tools for each of these processes was relatively easier than it is today. Back then, big data was catching on, with enterprises experimenting, and related technologies were also in their infancy. Today, the situation is drastically different. Data integration is no longer confined to only ETL (extract, transform, load) and real-time data replication but now encompasses stream processing and data virtualization, which find their way into all data architectures. Moreover, data lakes and NoSQL database implementations are on the rise, while relational database management systems (RDBMSs) and enterprise data warehouses are moving to the cloud.

On the data delivery end, data discovery, data preparation, and advanced visualizations are complementing traditional reporting and laying the foundation for data democratization. On the data governance front, metadata management, data cataloging, data provenance, and business glossary requirements are also gaining significant attention. Finally, some data platforms now promise to provide most of these capabilities (if not all) under a single umbrella.

What makes the task of choosing the right technology significantly more difficult is the plethora of technologies on the market, sufficient to perplex even a seasoned data architect. A decade ago, there was a choice of between nine to 15 tools across all data areas (data integration, persistence, and delivery); today, that choice has expanded to anywhere from 30 to 50 options. The exhaustive feature list of various tools adds additional complexity to the decision-making process for the "right fit" technology selection.

The choice of tools and technologies is not a "make or break" decision for any data architecture using modern technology; most of the technologies on the market are equally adequate and come with their own advantages and limitations. A perfect technology stack is nothing more than an urban legend; however, having a pragmatic framework for choosing tools and technology is essential. Figure 1 shows a framework with some important dimensions that can help in the evaluation of tools and technologies.

Based on the specific enterprise context, each dimension in this framework must be weighted in order to arrive at a shortlist that can be evaluated in depth. The aim is to select the tools and technologies most suitable for the specific needs of each enterprise. The right technology needs an ecosystem where people are engaged and follow processes correctly.

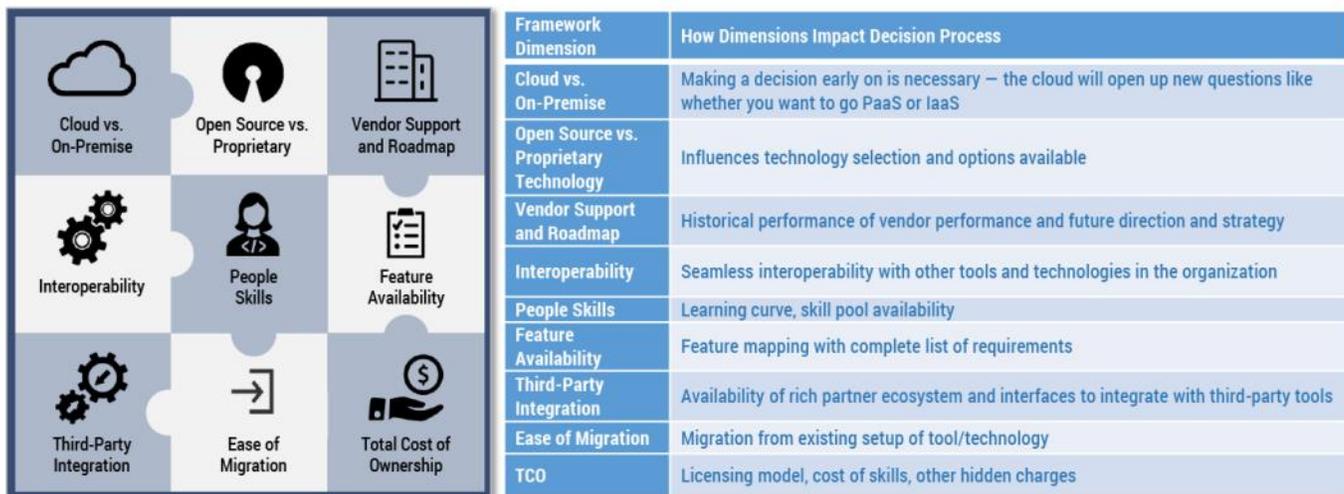


Figure 1 – Guidance framework for shortlisting technology components.

People and Process Transformation Influencing Data Architecture

Modern data governance tools can provide insights into how to enable and motivate stakeholders to collaborate and contribute to the data architecture. For years, data governance has faced challenges such as data architecture not being considered a shared responsibility but rather as the burden of the IT department. Other obstacles have surfaced in the form of regular maintenance issues, difficulties with extensible architecture, and data silos. To solve these challenges, modern data governance tools use crowdsourcing and gamification. These same concepts can be leveraged and contextualized to build a practical data architecture governance framework.

Today, any initiative is incomplete without an enabling technology platform, and data architecture governance is no exception. Crowdsourcing would imply that many data stakeholders will contribute to the data architecture, which may lead to chaos. A technology platform with built-in control tools (e.g., workflow approvals, checks for duplicates) will facilitate the structured collection and validation of information. The data architect, who directly collaborates with the data stakeholders to seek clarifications and additional information, is the final approver. Reports and visualizations built on top of the data show the various data systems, data sources, and data provenance. As additional information starts to trickle in, the data architecture diagrams will auto-update. As a start, as Figure 2 illustrates, the technology platform can capture the following six elements:

1. **Sources and targets** – provides an understanding of the external interfaces for the enterprise
2. **Data provenance and lineage** – offers a view of data flow across the enterprise
3. **Underlying tools and technologies** – helps identify possible technology rationalization opportunities
4. **Data trustworthiness** – provides a list of trustworthy data assets to which the business can refer
5. **Data criticality** – offers an understanding of what data is critical for the business and allocates resources to improve quality, timeliness, and so on
6. **Data sensitivity** – protects sensitive data and complies with regulations

The technology platform can employ gamification features such as points, badges, and levels. Contributors who capture (i.e., collect) the basic data elements listed above are rewarded with badges and awards. This increases participation and motivates stakeholders to be part of the journey. Socializing the leaderboard and success stories across the enterprise increases adoption. Moreover, the technology platform can be leveraged to define targeted goals. For example, data protection regulations, such as the European Union’s General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), are an important concern for enterprises, and identifying sensitive data is a top priority. The technology platform can focus on regulations and sensitive data and reward contributors

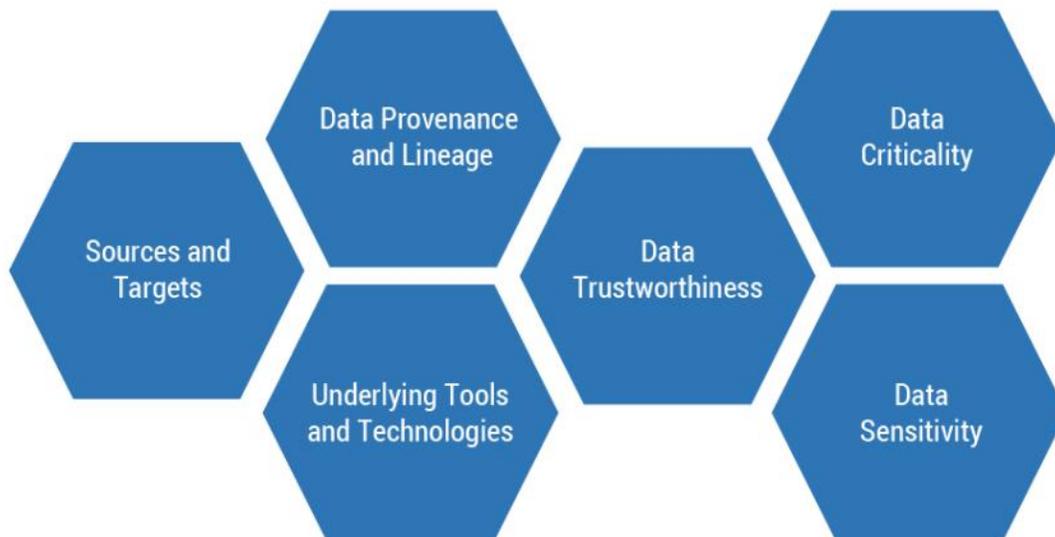


Figure 2 – Six elements of the technology platform for data architecture.

with special badges. Such initiatives ensure that the data architecture remains aligned with organizational goals. However, gamification alone will not ensure that data stakeholders remain active on the platform. They also need to be rewarded with some tangible benefits for investing their time and effort.

Addressing stakeholders' pain points ensures their adherence to data architecture management and governance.

In addition, dashboards and reports can be built on the platform to help data stakeholders:

- Identify trustworthy data sources/systems across the enterprise
- Better understand the impact on their data assets of upstream and downstream changes
- Recognize alternative data systems they can fall back on in the event of an outage in their systems, etc.

Addressing stakeholders' pain points ensures their adherence to data architecture management and governance.

Data architects play a vital role in the evolution of this technology platform. In Agile terminology, they play the role of product owners and ensure new features are added to the platform at regular intervals. For example, a chatbot feature can be added to address data stakeholders' questions. Data architects need to strike the right balance between addressing data stakeholder pain

points and gathering information to build and enhance the data architecture. Such an approach will help in creating a data architecture that:

- **Is easy to comprehend and socialize.** Reports generated out of the platform will be easy to understand and available to all data stakeholders.
- **Is easy to maintain.** The data architecture is built and managed collectively by all relevant stakeholders.
- **Addresses data silos.** Visibility across enterprise data assets provides the data architect with an opportunity to identify data silos and address them.
- **Prevents new data silos.** Visibility across enterprise data assets provides data stakeholders with the ability and opportunity to identify data assets that address their data needs without creating new data systems.

Conclusion

For decades, enterprise data architecture has been perceived as complex blueprints confined to the realm of the IT department. In reality, a majority of data assets are outside the purview of IT. Gaining the confidence of and onboarding various stakeholders, such as power users, data custodians, shadow IT departments, and data consumers, will help enrich the data architecture and provide benefits for everyone. On one hand, the data consumers and power users will have access to trustworthy data sources of which they previously may not have been aware. On the other hand, the IT department will have a better understanding of data

silos and can focus on reducing them. Moreover, shadow IT and data custodians will gain a better understanding of the impact on their data assets/systems of upstream and downstream changes and of the impact their assets/systems have upstream and downstream, thereby contributing to the purpose of shared vision and responsibility.

A mindset change supported by the harness of automation is the right sauce for success in the digital era. We can only realize the true manifestation of the power of digital technologies when we can realize the vision of an intelligent enterprise.

Reference

¹“Winning in a Business 4.0 World.” Tata Consultancy Services, 2018.

Sagar Gole is Business Intelligence Solution Architect of the Automotive & Industrial Business unit of the Manufacturing & Utilities Business Group for Tata Consultancy Services (TCS). For the last 16 years, he has engaged in designing and building business intelligence, data integration, and data management solutions. In his current role, Mr. Gole advises customers on building data platforms (enterprise data warehouse and data lakes) via consulting. He can be reached at sagar.gole1@tcs.com.

Vidyasagar Uddagiri is Enterprise Architecture Practice Leader of the Automotive & Industrial Business unit of the Manufacturing & Utilities Business Group for Tata Consultancy Services (TCS). Over 23 years, he has worked in application development/maintenance, technology modernization, and enterprise architecture in the manufacturing, banking, and retail domains across North America, Europe, and the UK. In his current role, Mr. Uddagiri’s key responsibilities involve business development via consulting and digital transformation strategies definition and technology evangelization for customers by means of creating point of views and proof of concepts. He can be reached at vidyasagar.uddagiri@tcs.com.

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

"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

"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

"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

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.