



Data Insight
& Social BI

Vol. 13, No. 5

Data Quality and Governance in Projects: Knowledge in Action

by Danette McGilvray and Masha Bykin

Organizations make substantial investments in projects to achieve business goals and seize opportunities. The more effective those projects, the sooner they realize results. Increase the speed and success of your projects by making data quality and data governance activities an integral part of the solution (or software or systems) development lifecycle, or SDLC. This *Executive Report* discusses the critical activities needed for both sequential and agile methodologies. Including these will also improve the quality of the data that matters most, transforming it into a valuable asset.

Report

Access to the Experts

Cutter Consortium, a unique IT advisory firm, comprises a group of more than 100 internationally recognized experts who have come together to offer research, consulting, training, and executive education. These experts are committed to delivering top-level, critical, and objective advice. They have done, and are doing, groundbreaking work in organizations worldwide, helping companies deal with issues in the core areas of software development and agile project management, enterprise and business architecture, business and technology trends and strategies, innovation, enterprise risk management, metrics, and sourcing.

Cutter offers a different value proposition than other IT research firms: We give you *Access to the Experts*. You get practitioners' points of view, derived from hands-on experience with the same critical issues you are facing, not the perspective of a desk-bound analyst who can only make predictions and observations on what's happening in the marketplace. With Cutter Consortium, you get the best practices and lessons learned from the world's leading experts — experts who are implementing these techniques at companies like yours right now.

You can tap into this expertise via print and online research services and journals, mentoring, workshops, training, and consulting. And by customizing our information products, training, and consulting services, you get the solutions you need while staying within your budget.

Cutter Consortium's philosophy is that there is no single right solution for all enterprises, or all departments within one enterprise, or even all projects within a department. Cutter believes that the complexity of the business technology issues confronting corporations today demands multiple detailed perspectives from which a company can view its opportunities and risks in order to make the right strategic and tactical decisions. The simplistic pronouncements other analyst firms make do not take into account the unique situation of each organization. This is another reason we present the several sides to each issue: so you can determine the course of action that best fits your unique situation.

Expert Consultants

Cutter Consortium products and services are provided by the top thinkers in IT today — a distinguished group of internationally recognized experts committed to providing top-level, critical, objective advice. They create all the written deliverables and perform all the consulting. That's why we say Cutter Consortium gives you *Access to the Experts*.

**For more information, contact
Cutter Consortium at +1 781 648 8700
or sales@cutter.com.**



The Data Insight & Social BI Executive Report is published by the Cutter Consortium, 37 Broadway, Suite 1, Arlington, MA 02474-5552, USA. Tel: +1 781 648 8700; Fax: +1 781 648 8707; Email: service@cutter.com; Website: www.cutter.com; Twitter: @cuttertweets; Facebook: Cutter Consortium. Group Publisher: Chris Generali, Email: cgeneral@cutter.com. Managing Editor: Cindy Swain, Email: cswain@cutter.com. Print ISSN: 1540-7403 (Executive Report, Executive Summary, and Executive Update); online/electronic ISSN: 1554-7078.

©2013 Cutter Consortium. All rights reserved. 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 more information about reprints and/or back issues of Cutter Consortium publications, call +1 781 648 8700 or email service@cutter.com.



Rob Austin



Ron Blitstein



Tom DeMarco



Lynne Ellyn



Israel Gat



Vince Kellen



Tim Lister



Lou Mazzucchelli



Ken Orr



Robert Scott

Cutter Business Technology Council

Data Quality and Governance in Projects:

Knowledge in Action

THIS ISSUE'S AUTHORS



Danette McGilvray



Masha Bykin

IN THIS ISSUE

- 1 Real Challenges
- 2 Organizations, Data, and Projects
- 3 Why We Overlook Data
- 3 Myths and the Reality of Projects
- 5 Data Knowledge Network
- 7 Data Quality Techniques
- 7 SDLC with Data Governance and Quality
- 14 Agile with Data Governance and Quality
- 15 Best Practices for Collaboration
- 17 Successful Projects
- 18 Conclusion
- 19 Endnotes
- 19 About the Authors

Organizations invest substantial resources in projects. From exciting new innovations to mundane yet necessary improvements, companies institute projects to develop processes and practices that help the business address issues, support goals and objectives, or take advantage of opportunities. Clearly, the more effective the projects, the less cost to the company and the sooner the business can use those results to provide products, offer services, and increase revenue.

Historically, many projects concentrated their efforts on people, process, and technology. However, a vast amount of projects still fail to fully address the data and information aspects of their efforts. Many projects have failed due to this oversight, while others have left a wake of data quality issues that put long-term burdens on business processes and subsequent projects. We can increase the success of our project portfolios by making data quality and governance activities an integral part of our project approach.

The structured approach to delivering these projects is called the solution (or software or systems) development lifecycle, or SDLC. This *Executive Report* discusses critical activities related to information quality and data governance in every phase of a sequential-type SDLC, mapping to an agile Scrum methodology, and why these activities are important to include. The reader will walk away with a list of concrete activities to incorporate into any SDLC and the ability to articulate the value of their inclusion with project managers and stakeholders.

REAL CHALLENGES

To begin, let's look at ACME, a fictitious company facing real challenges. Improving the online customer experience is one of the top strategies of ACME, a manufacturer and distributor of health products that sells directly to consumers and resellers such as grocery stores and sports clubs. Bringing robust sales and service features to the Web for self-service is the first priority of Patel, a senior-level operational manager. The current website lacks integration with internal systems

and processes. As such, functionality has been duplicated between applications that serve different channels, resulting in process inconsistencies and data issues that frustrate both customers and internal staff. The company needs cross-functional, integrated services and data. Patel plans to modify and replace capabilities gradually with new components that support both public and internal use. Architectural improvements — such as centralized Web services, customer identity resolution, and product mastering for quoting, tracking orders, and billing — seem necessary to achieve these goals.

Organizations invest a substantial amount of resources and time in projects. In spite of this, many projects produce less than stellar results.

So Patel initiates a project to measure online sales and service results, site usage, and feedback using business intelligence and analytics techniques. BI uses the data warehouse, some of which is virtual. The warehouse consolidates operational data across the company and has been in production for almost five years. Several business units already rely on BI reports that analysts create out of the data warehouse and the virtual BI environments, some of which are also used for government-mandated reporting and audits. Patel heard there are many conflicts in different reports that are asking the same questions, and he's not sure how to resolve these issues. The lack of consistency across reports causes declining trust in the data and information systems. Unless he can address the data quality issues, this problem could impede progress on the operational capability improvements needed to enhance the experience of the companies and consumers who purchase from ACME.

ACME's executive leadership team has also prioritized the need for integrated functionality for manufacturing, inventory and order management, sales tracking, and finance. The problem: the company spends too much money on standalone legacy systems. Some capabilities are missing or some amount of integration is not possible, frequently due to limitations of the technology itself. Many of the systems have become brittle after years of workarounds and there are complaints of "bad data." The bottom line is that the legacy systems will not get the company to where it needs to go. The plan includes implementing an ERP application and migrating data from standalone systems. The project is well underway and Cynthia, a senior-level manager, is

accountable for the success of the overall program — from technology implementation to business process changes to organizational impacts and acceptance.

These situations are all too common, and you might face similar challenges. Does this scenario sound familiar? You begin your project. You choose the necessary tools and vendors, sign the contracts, select the project methodology, assign business and technology subject matter experts (SMEs), approve the budget, and set the timeline. You allocate every aspect of people, processes, and technology into the plan. But what about data and information? Have you accounted for this in the plans? By the time someone notices such an oversight, you have already hired programmers and consultants and put them on payroll. Management, including yourself, feels the pressure to put these expensive resources to work, writing code to move and integrate the data, even though nobody knows what the data actually looks like and only high-level reporting requirements are known.

Every day, around the world, organizations¹ (for profit, nonprofit, healthcare, education, government) struggle with similar problems. They invest a substantial amount of resources and time in projects. In spite of this, many projects produce less than stellar results. Through this *Executive Report*, we want to inspire project teams to improve and protect performance in the marketplace and to be vehicles of positive change.

ORGANIZATIONS, DATA, AND PROJECTS

An important result of your project should be information² — information you can trust in the months and years following the implementation. The outcome should not be just a better process or system, but better *information* — allowing executives to make informed decisions and take effective action. However, in your meetings, what is the typical topic of conversation? Usually which application you will purchase and who is the vendor. Those are important issues, but equally important considerations are how your company will use the information to conduct business, the readiness of existing data, and how you can adapt the data to fulfill the requirements of the new system. Just as a conscious focus on processes, people, and technology are essential and usually planned for, the data quality and governance aspects require the same level of attention.

Earlier we introduced the SDLC. This is the structured approach to delivering projects. Various SDLCs used by IT departments typically focus on technology implementation (replacing System X with System Y), or the delivery of new functionality (screens) or processes (use cases). Even the most comprehensive SDLCs frequently

miss the right level of focus on an important component — the data and information. If your projects do not explicitly address data, then the information needs they appear to promise will eventually prove to be absent. The situation will manifest as mistrust of the new technology and complaints of “bad data,” and the many root causes will be difficult to correct. These problems feed the cycle of underperforming systems and replacement projects, which hinder any company’s ability to fulfill its mission. Your organization can do better; you already have the ingredients, and this report will help.

Let’s move ahead by answering a question not often asked, “Why is data often forgotten in our projects?”

WHY WE OVERLOOK DATA

Many companies are still waking up to the importance of data and the need to manage it more closely within projects. There are a couple of reasons for this: (1) we take the existence of information for granted (i.e., it will be there when we need it); and (2) we passively assume that someone else is taking care of it (i.e., records are updated, reports are run, and dashboards are refreshed according to schedule).

The often unspoken assumption is that the information is “good,” which is another way of saying that it has good data quality. What does “good” information mean? It means that:

- We can *find* what we need (we can get to it; we can access it).
- It is available *when* we want it (it is timely; it is not late).
- It includes *everything* we need (nothing is missing).
- We *understand* it when we see it (we can interpret it).
- It is *correct* (it is an accurate reflection of what’s happening or what did happen in the real world).

Because of all these things ...

- We *trust* it when we get it.
- We can actually *use* it with confidence.

Information and data are behind most company decisions. That the information is “good” is a big assumption. What happens if that assumption fails us? In the case of ACME, managers hear complaints that the reports are wrong (incorrect), they don’t contain the information needed (something is missing), or the reports are arriving too late for anyone to take action (not timely). People show up at meetings with differing answers to the same questions — ready to fight to

defend reports that prove their case but can’t be reconciled to other reports. They rely on custom spreadsheets and spend time on analysis that has been duplicated by multiple teams and repeated several times a year, with different outcomes. People argue and waste time and money. All this effort and still people make uninformed decisions or delay action due to uncertainty.

On the operational side, order systems show that products are in stock when they are actually on backorder. Customers complain about not being informed when placing their orders. Morale among customer service reps is low since they get the brunt of customer dissatisfaction. Some customers have even taken their complaints to Facebook and Twitter. In this age of real time and social media, unhappy customers and poor decision making are threats to even the most beloved brands.

MYTHS AND THE REALITY OF PROJECTS

It is useful to understand some of the common yet erroneous beliefs associated with projects. Believing such myths has caused many projects to fail because project teams did not account for the reality in their plans and budget. Let’s examine some of these myths and realities, along with their solutions.

Myth: The software implementation is a technology project.

Reality: The implementation is a business project that requires partnership with your technology group. The implementation will change how you conduct business, it will change your business processes, and therefore impact your people and the data that supports the processes.

Solution: A successful implementation requires the involvement and collaboration of those with expertise in business, technology, and data. It also includes attention to communication, training, and helping the organization move through and adopt the changes caused by the project implementation.

Myth: We don’t need to worry about data quality; our application will guarantee accuracy.

Reality: While you can build an application with data quality in mind, the following are examples of data quality issues that the application alone cannot fix:

- **Data entry issues and subjective choices.** Information entered into a field may vary from person to person. It is not unusual to have four lines of free-form text as part of an address, with one line indicating the division or department. Some will put the division or department in the first address line, while

others will use the second or last line. A data element may require a decision about the association of a product to a product category, such as “hoodie sweatshirt” to “tops” or “jackets.” This decision is subjective, and people can choose differently. Another scenario is that people can be misinformed, and thus make poor choices. For example, a field requires a data entry clerk to designate whether a part in inventory is made inhouse or outsourced. If a business process doesn’t clearly mark boxes with this information, the clerk may indicate the wrong value.

- **Incorrect mappings during migration/conversion.** A system may allow us to restrict a data element to five values. During migration, values from source system records can be mapped incorrectly or loaded as “unknown.” For instance, a column defined as “wholesale price” can be mapped incorrectly into “retail price,” even if rules are in place to restrict formats and value ranges.
- **Limitations in the technology.** If a system does not have the ability to implement an important rule or process, users may be required to do cumbersome workarounds that are prone to error. Lack of interoperability between systems can cause issues that require manual corrections after the data moves between systems.
- **Data duplication.** Creation of duplicate records can be hard to control with application interfaces. A common example is customer records. If people don’t perform the task of searching for existing records, and proceed to create new ones instead, they will create duplicates. This is one of the most common and most difficult data quality issues to address after systems are in place for months or years.

Solution: It is true that the software application can help improve data quality by implementing data quality rules, and by displaying data definitions and other context-specific help content to the user. However, the rules and the help content must be defined by people who recognize the attributes of good and bad data. Some of the rules are themselves requirements that need to be defined, which must be developed and tested to ensure good data quality.

High-quality data depends on execution of stewardship business processes by trained people who are motivated and rewarded to ensure data quality. Once we can define the data quality rules, the technology can do a lot that will positively impact data quality, but technology cannot figure out the rules for us. If we bring the practice of data quality analysis and stewardship into projects, we treat key data assets with appropriate care, find

opportunities to make systems more friendly, and identify the many needs around training and documentation, avoiding costly and disruptive problems.

Myth: We are under a tight timeline. We don’t have time to deal with data quality. We’ll just load the data and clean it up later.

Reality: If you don’t pay attention to your data quality, chances are you won’t be able to load the data in the first place. And if by chance you *can* load the data, your ability to conduct business without interruption is at risk.

Solution: Dealing with data quality early helps you *keep* to your timeline, not miss your deadline. See the next myth.

Myth: All the data we need for our new application currently exists in legacy systems. We just need to find it and simply move it into the new system.

Reality: Any statements about the existing data are assumptions until you can view the actual data and compare it to the new requirements. You cannot fully know the gaps between the current state of data and requirements during initial planning. Data profiling and the involvement of data stewards provides input to this gap analysis based on the actual data. There are three outcomes from the comparison:

1. Existing data meets the new requirements, so no changes are needed.
2. Existing data needs to be changed in some way to meet requirements, or existing data is being used in a new way. This often means cleansing, correcting, or enhancing the data, preferably at the source. It can also require writing code to transform or integrate the data during migration.
3. Data needed for the new application does not exist in the legacy systems, or the quality is so poor that it will not meet the new requirements, which means you will have to create data. In some cases, it is appropriate to purchase data to fill a gap. However, this data must itself be assessed for quality and then integrated correctly into inhouse data.

In this report, we refer to “data readiness” as the resolutions that close the gap between current and required data quality (see sidebar “Data Readiness”).

Solution: The earlier in the project you start your data readiness activities the better. The sooner you discover the gaps between the actual and required data quality, the more quickly you can account for resources needed to best close those gaps, and the more time you have to implement solutions. Realize that creating data from

scratch will not ensure good quality unless you identify and document requirements, control the processes, train those creating the data, and test the data before migrating it.

Myth: We don't need to worry about the data until we do our testing. We will find anything we need to know at that time.

Reality: Often data ETL (extract, transform, load) programs are written based on assumptions about the data, which do not account for the actual state of the data (e.g., column heading of "Phone No." for a field that actually contains ID numbers).

Solution: Avoid the scenario of "code-load-explode" by assessing your data readiness early in the project. Issues found early are easier to fix. The same issues found during testing are more costly to fix and have a greater delay to the timeline. For example, it is commonly accepted that requirements defects found during testing cost 10 times more to fix at that time than whatever it would cost if they were addressed in the requirements phase when first discovered. If they are found after deployment, they can be 100 times more costly to fix.³

Myth: It's just business as usual. We can continue the same practices, using the same mindset, as we have in the past and the project implementation will be successful.

Reality: A new level of data coordination, cooperation, communication, and quality is needed for success, because business is expected to adapt to the Information Age. Consumers and business partners demand more intelligent use of data and transparency in information exchange. Project teams require the same.

Solution: This is where data governance, stewardship, and data quality activities apply.

DATA KNOWLEDGE NETWORK

It's not easy to get the right people together to make decisions and take effective action regarding data. In our complex operational and project environments, it can sometimes feel impossible. It is important to put into place a process where knowledgeable people representing data, business, and technology together can raise and prioritize data issues, make decisions, implement changes, and communicate. The following three data management components will enable companies to put such processes into place:

1. Data governance — defines and enforces rules of engagement, decision rights, and accountabilities for information assets by organizing and implementing

DATA READINESS

Data readiness is the process of ensuring that the data within a project scope is prepared to meet new business requirements. This means closing the gap between the current state and required state of the data; for example, through data clean up, business process improvements in legacy source systems, writing transformation code, or creating higher-quality data.

the right level of policies and procedures, structure, roles, and responsibilities.⁴ Issues surrounding data quality can come from anywhere — project team members, users, management, business, and IT. We use the phrase "data stewards" to refer to those people who have decision rights. Depending on the size and organizational model of a company, there may be several kinds of data stewards with varying degrees of accountability, and varying degrees of formal process. The "right-sized" governance for your organization provides venues for interaction and communication channels through which you can discuss data-related issues. Data governance is often implemented as an independent organizational team that oversees data management activities. They coordinate the activities of the data stewards, who are often on other teams that use data.

- 2. Data stewardship** — the formal or informal accountability for business (nontechnical) control and use of data assets.⁵ Data stewards perform the activities of stewardship, such as establishing data requirements, definitions, quality rules, and measures. Data stewards are the SMEs who research issues and make decisions on matters impacting their domain. They bring in other SMEs as needed and make use of governance processes for cross-functional issues.
- 3. Data quality** — the degree to which information and data can be a trusted source for required uses.⁶ In this report, the phrase "data quality" is also used to refer to the activities that ensure the quality of information and data. Data quality activities include data readiness assessments, evaluating data contents against requirements, and determining how to close the gap between actual and required data. Data quality activities also include root-cause analysis, ensuring the consideration of both preventive and corrective measures and enabling timely discovery, notification, and response to issues. A team of data analysts with decision-making guidance from data stewards frequently performs this work.

The allocation of resources toward data governance, stewardship, and quality activities in an organization creates a network of people who may not otherwise interact, be aware of each other's skills and expertise, or have clear decision rights. The combination of data governance, stewardship, and quality activities and related resources, along with their communication channels, is the "data knowledge network." This network will naturally include others in the organization who normally participate in data management, such as data modelers, database administrators, architects, and developers.

If your organization already has formal data governance in place, use it in your project; it will help empower and strengthen the existing network. Ensure your project activities and resources involve those with expertise in data. Leverage the communication paths established by data governance to ensure appropriate representation, decisions, identification, and resolution of issues. Depending on how your project is set up, designated data stewards may be on the core team or part of an extended set of resources used as needed throughout the project.

If your organization does not have formal data governance, stewardship, and quality functions in place, then discover the informal data knowledge network that is already in place, as all organizations have them. Determine if you can make use of existing processes, expertise, and decision-making bodies in your organization to carry out the work described above. Establish a decision-making process for data issues for your project. Attempt to make it somewhat independent of your project, so that it can remain in place after your project closes. Production operations and future projects can then leverage the same processes, evolve the network, address issues, and sustain improvements in data quality.

Maturing Professions

The structured professions and disciplines of data governance and data quality have matured a great deal in the last 10 years. It is outside the scope of this report to discuss how to set up formal governance, stewardship, and data quality processes and functions. However, there are many resources available for learning more and evolving your practice. Professional groups like DAMA International, IAIDQ (International Association for Information and Data Quality), and DGPO (Data Governance Professionals Organization) sponsor conferences and provide resources like the *Data Management Body of Knowledge (DMBOK)* and certification such as the Information Quality Certified Professional (IQCP) and the Certified Data Management Professional

(CDMP). Moreover, the University of Arkansas at Little Rock offers advanced degrees in information quality through its Information Quality (IQ) Graduate Program, which was established in collaboration with MIT's Information Quality (MITIQ) Program. Several countries (France, Germany, the UK, Australia, South Africa, and Brazil, among others) also have their own professional associations, chapters, and/or conferences devoted to data quality and governance. Experienced consultants, including those from Cutter Consortium, are available to help organizations address the data quality and governance aspects of their business efforts.

Yet, if you limit data governance and quality activities to just data quality projects, such as data cleanups, then you will miss the many opportunities to improve organizational data quality. The quantity of data quality problems will continue to increase and compete for resources with more value-adding work within the company. It is the deliberate and consistent inclusion of these activities in the SDLC that achieves the creation of quality data, enabling investment afterward in value-adding work instead of rework.

Including the Data Knowledge Network in Project Work

It is critical for your project that data stewards, those with decision rights to represent the various areas impacted by the issues, are part of the discussion, decisions, and action. Looping in the appropriate parties to make decisions is a core capability of the data knowledge network. The simplified data-to-process matrix in Table 1 shows how decisions about data usually span more than one functional area. A similar reference, at the right level of detail, can be an artifact created and maintained by a data governance or project team, to assist with identifying the right resources.

If there is a change to, or issue with, a business process, members of the project team can look down the column to see which data elements will be impacted and include the appropriate steward(s). If there is a change to, or issue with, the data, look across that row to determine possible impacted business processes and include the associated business representatives. Be sure to involve technology teams to represent the related applications and systems for impact assessment, architecture, root-cause analysis, and many other discussions about changes.

Another way to loop in the network is to establish a dedicated consultative role on the project team for a "project data steward." This individual would participate in project activities that are likely to generate issues such as requirements, solution designs, and data analysis. They would be responsible for monitoring project

Table 1 — Data-to-Process Matrix: Identifying Shared Responsibilities

	Key Business Processes				
	Quote to Cash	Procure to Pay	Manufacturing	Product Lifecycle	Finance
Master Data					
Item	●	●	●	○	●
Customer	○				●
Price List	●	○		●	○
Supplier			●	●	●
Etc.					
Transactional Data					
Sales Orders	○		●		●
A/R Invoices	○		●		●
Etc.					
○ = process creates the data ● = process uses the data					

©2005 Danette McGilvray, Granite Falls Consulting, Inc.

artifacts and activities for possible gaps in definitions and data quality issues, coordinating the participation between the project team and necessary data stewards, and aligning the project with other work in progress.

DATA QUALITY TECHNIQUES

Data quality dimensions are aspects or features of the data. They provide a way to measure and manage the quality of data and information.⁷ Examples of data quality dimensions include accuracy, timeliness, consistency, and synchronization. Each data quality dimension requires different tools, techniques, and processes to measure and manage it, resulting in varying levels of time, money, and human resources. By understanding which dimensions are applicable to the goals of the project, you can better scope and plan. Initial assessments of the chosen data quality dimensions set a baseline for your project and identify issues you need to address. Informed actions addressing the root causes of the issue — and cleansing or transforming data as needed — can be built into your project plan. Continuing assessments provide visibility to progress and indicate the readiness of your data for the move to production.

Data profiling is the use of analytical techniques to discover the structure, content, and quality of data.⁸ You can perform profiling using various technical means (e.g., writing SQL or using one of the data profiling tools available on the market). Profiling is *not* equal to having data quality. It is, however, one of the first and most basic techniques to discover the existence, structure, content, validity, and quality of data. Most other dimensions of data quality, such as

accuracy and duplication, build on what is learned from profiling data.

The category of tools most commonly referred to as “data quality tools” streamline or automate data profiling, monitoring, and/or cleansing activities. Other tools that support data quality analysis include data modeling, glossary and taxonomy, ETL, and metadata collection and exchange. The activities, content, and artifacts needed for high-quality data can be very time-consuming to perform manually or to maintain as documents. Tools can bring enormous efficiencies. However, if there is no clear direction based on sound governance and stewardship practices, tools can distract (i.e., thousands of features and options) and create a false sense of security (i.e., “something” is being done). Introduce tools only when you know how they will add value to projects.

SDLC WITH DATA GOVERNANCE AND QUALITY

Projects trigger changes that raise issues and require many decisions. When not informed by the data knowledge network, unintended consequences often ensue. When we do incorporate data governance, stewardship, and quality activities, we trigger the project to pay attention to data and mobilize the data knowledge network to help. From this, we get optimal just-in-time decisions about data made collaboratively by those who need to be involved. Issues get detected and raised at the appropriate time, allowing for proper solutions for work in progress and the avoidance of new, costly issues. Data quality analysis informs the response to issues raised, while the knowledge of data stewards informs the solution options. Project team members can

then make better decisions, such as which issues take priority and how to effectively correct root causes.

Once the importance of paying attention to a project's data is clear, the immediate question is "How do we include a focus on data in our SDLC?" Next, the report outlines seven phases — initiation, planning, requirements and analysis, design, build and test, deploy, and production support — to illustrate integrating data quality throughout the SDLC.

SDLC Phases

The following are the high-level statements of what is included in each phase:

- **Initiation** — articulate the problem or opportunity to be addressed; put the project into motion; authorize and define scope and goals; and initialize resource allocation
- **Planning** — refine project scope and goals; develop a plan for managing, executing, and monitoring the project; identify project activities, dependencies, and constraints; and develop initial timeline
- **Requirements and analysis** — research, evaluate, and identify detailed needs and objectives; establish priorities; and sequence activities and refine plans
- **Design** — define solution options and select best option based on dependencies and constraints;

determine how to meet the functional and quality requirements of the solution; and adjust plans and requirements as needed

- **Build and test** — construct the solution and determine if requirements have been met; and adjust plans, requirements, and design as needed
- **Deploy** — move the solution to production; deliver to users; stabilize the deployment; and transition to operations and support
- **Production support** — maintain and enhance the solution; and support users in the operational environment

Several SDLC models are in use to manage the complexity of modifying or creating systems, software, and solutions. Phases vary in name, organization, coordination, timing, and formality, depending on methodology and other factors determined by project teams. Though the names and number of phases in SDLCs will differ, you can map most to the seven phases used here. For example, in almost every project there is some type of activity related to gathering requirements. You may call that phase by a different name; however, if you are gathering requirements, some data quality activities should be included during that phase of your project.⁹ No matter which SDLC you apply, you will be able to see where the activities fit.

Table 2 — SDLC Phase: Initiation

Initiation Phase Team Activity	Data Governance and Stewardship Activities	Data Quality and Readiness Activities
Define business problem and/or opportunity; establish scope; set goals.	Identify data subject areas required for scope and goals (e.g., customer, order history, products). Set goals for data quality within the context of overall scope and goals. Evaluate availability of definitions for data elements within the domains. Articulate how high-quality data and information support business goals and poor-quality data hinders them.	Identify possible sources of data within scope. Collect known data quality issues, existing quality metrics, and assess trust. Identify potential risk and impact to the project of data quality issues found.
Begin initial resource allocation.	Ensure data stewardship and governance activities are accounted for when negotiating contracts, allocating human resources, approving budgets, and setting timelines. Assign resources for initial planning, requirements analysis, and to support data quality assessments.	Ensure data quality issues, activities, and tools are accounted for when negotiating contracts, allocating human resources, approving budgets, and setting timelines. Assign resources for initial data quality assessments and to support requirements analysis.

©2013 Danette McGilvray and Masha Bykin

SDLC Activities

Tables 2-8 provide an overview of activities related to data governance and quality in the SDLC phases just described. The first column indicates the SDLC phase and typical activities that a project team would be engaged in during the phase. The next two columns list high-level data governance, stewardship, and quality

activities during that team activity. We have grouped data governance and stewardship together because they are highly interdependent, nontechnical, and focus on decision making, whereas data quality activities focus more on the skills, techniques, and analysis that feed into solutions and decisions.

Table 3 — SDLC Phase: Planning

Planning Phase Team Activity	Data Governance and Stewardship Activities	Data Quality and Readiness Activities
Determine how to manage and monitor the project.	Determine how data governance and stewardship will engage with the project team. Plan for tracking and reporting status of data governance and stewardship activities. Determine how data knowledge network will interact with project and nonproject resources for data readiness activities.	Determine how data quality resources will engage with project team. Plan for tracking and reporting status of data quality activities.
Research activities that will support goals.	Determine existence and completeness of glossary, data models, and other metadata for data sources; identify gaps and necessary activities to close the gaps. Identify the desired population of the data subject areas at a high level (e.g., all active customer records, order history from past 10 years, all current and obsolete products from past five years).	Conduct quick, high-level data profiling of main data sources. Use to provide input to selection of data sources and initial insight into data quality issues that need to be accounted for during the project and in this planning phase. Help assess impact/risk to the project of data quality issues known at this time.
Identify high-level activities, dependencies, and constraints.	Identify needed data governance and stewardship activities and incorporate into project plan. Reference activities from SDLC phases. Prioritize known data quality issues and work with data quality analysts to plan solutions. Track progress of data readiness activities and manage as a dependency.	Identify needed data quality and readiness activities and incorporate into project plan (e.g., data profiling and other assessments, plan solutions to data quality issues). Reference activities from SDLC phases. Identify dependencies and constraints that affect or hinder ability to carry out data quality activities and help to assess impact/risk to project. Work with nonproject teams to start data readiness based on known issues and current business needs (e.g., good housekeeping corrections, reduce data volume), not based on new requirements.
Develop initial timeline.	Estimate time and resources to perform stewardship to achieve goals. Allocate time for analysis of data issues and decisions.	Estimate time and resources to perform data quality activities to achieve goals, including known data quality issues.
Fine-tune resource allocation.	Adjust resource allocations to account for data governance and stewardship activities identified during planning.	Adjust resource allocations to account for data quality issues and data readiness activities identified.

©2013 Danette McGilvray and Masha Bykin

Table 4 — SDLC Phase: Requirements and Analysis

Requirements and Analysis Phase Team Activity	Data Governance and Stewardship Activities	Data Quality and Readiness Activities
Create use cases.	Identify data elements within use cases. Ensure definitions are consistent across use cases and are captured in a glossary.	Analyze the use cases to inform the information/data lifecycle (create, store and share, maintain, apply, dispose). (Source: McGilvray, pp. 23-30.)
Perform functional requirements analysis.	<p>Include data stewards in functional requirements analysis (analysis sessions, process reviews, and artifact reviews).</p> <p>Identify data elements associated with all functional requirements.</p> <p>Ensure business rules and data definitions and valid value sets are validated, aligned, and documented.</p> <p>Ensure alignment and updating of glossary and data definitions that will need to be used throughout the SDLC (design, build, etc.).</p> <p>Ensure requirements are informed by data quality assessments and data readiness activities. Include recommended improvements based on root-cause analysis. Ensure solutions consider both corrective and preventive measures (e.g., business process improvement, training, changes to roles/responsibilities, automated business rules, data quality monitoring).</p>	<p>Gather requirements for data quality dimensions such as integrity, completeness, timeliness, consistency, accuracy, de-duplication, etc.</p> <p>Use business rule analysis to ensure requirements for data quality measures are understood and documented — for testing, initial loads, and ongoing quality checks (to be done in production).</p>
Perform physical data analysis.	<p>Follow up on questions and issues uncovered by data quality activities.</p> <p>Determine impact of data issues and readiness gaps to project. Prioritize and add resolutions to overall requirements artifacts to ensure they are addressed during subsequent SDLC phases.</p> <p>Update glossary based on findings from data quality activities (e.g., business rules, calculations, valid values).</p> <p>Continue to work with nonproject teams to address data readiness issues and dependencies.</p>	<p>Perform in-depth data profiling and other applicable assessments (using full data sets) to identify gaps between the actual data and known requirements. Learnings will be used throughout the SDLC in design, build, and test.</p> <p>Finalize the data population of interest (i.e., the selection criteria) and how the data can be accessed.</p> <p>Ensure findings from assessments are reflected in requirements and artifacts to be handled during design, build, test.</p> <p>Identify root causes of data quality issues. Use as input into both requirements and design.</p>

©2013 Danette McGilvray and Masha Bykin

Table 5 — SDLC Phase: Design

Design Phase Team Activity	Data Governance and Stewardship Activities	Data Quality and Readiness Activities
Consider architecture.	Determine approaches to handling data based on requirements and identify gaps in existing architecture or tools (e.g., load using ETL tool or flat files).	Consider appropriate handling of all categories of data in high-level design (e.g., transactional, master, reference, configuration, purchased, and metadata).
Perform issue tracking and resolution.	Determine status of data readiness dependencies previously identified. Design workarounds for data quality issues that have not been resolved and that may need to be absorbed into the project. Prioritize and make decisions about solution designs.	Recommend solutions to close gaps between actual and required data quality, including both corrective (e.g., cleanse, correct, enhance, or create data) and preventive approaches.
Create data model design.	Participate in model creation. Align data models and glossary definitions. Identify interdependencies between subject areas and data elements.	Use data profiling results to provide input into the creation of target models. Consider differences between existing data sources and new target models to facilitate integration.
Create user interface design.	Provide input into interface design to balance ease of use and timeliness of task execution with features that protect the quality of data. Identify data elements in screens. Ensure definitions are consistent across screens and are captured in a glossary. Incorporate definitions into training and help content.	Use profiling results to expose valid value sets, rules, and data quality issues that could be addressed in the user interface.
Create data movement/ ETL designs.	Review and validate source-to-target mappings and transformation rules. Look for opportunities to standardize valid value sets and hierarchies. Ensure there is agreement on the overall flow of data from sources to targets.	Use profiling and other assessment results to identify data element contents so source-to-target mappings and transformations are based on the actual data, not column headings or opinions of content. Help determine the optimal order in which to load data.
Create test plans.	Identify and prioritize key data elements, measures, and rules to be validated during testing. Identify and prioritize key data elements, measures, and rules for ongoing monitoring (i.e., after go-live).	Identify sources for test data and profile to ensure contents are well known and comparisons will work (i.e., reduce time spent chasing issues during testing that are problems/deltas in the test data). Plan for testing for all categories of data within scope (e.g., transactional, master, reference, configuration, purchased, created data, and metadata). Assist in designing test approaches for data quality. Consider reusability so tests can be repeated during test cycles and after go-live.
Create deployment plans.	Ensure data artifacts and documentation are available (e.g., business rules, glossary) for the production of help, training, and communication content.	Document issues and solutions that involve manual processes and should be included in training and communication. Include data readiness steps that cannot be automated and must be performed manually before the system is released to the users. Include data quality checks in deployment plans (e.g., confirm correct data sources prior to load, data quality checks after).

©2013 Danette McGilvray and Masha Bykin

Table 6 — SDLC Phase: Build and Test

Build and Test Phase Team Activity	Data Governance and Stewardship Activities	Data Quality and Readiness Activities
Build features.	Support developers and analysts in researching questions about meaning and correct use of data.	Implement data readiness solutions and adjust based on requirements and design. Create new data as part of data readiness if needed. Implement business process improvements to prevent data quality problems after go-live.
Refine requirements and designs.	See Tables 4 and 5 for activities as adjustments are made to requirements and designs.	See Tables 4 and 5 for activities as adjustments are made to requirements and designs.
Perform issue tracking and resolution.	Ensure recommended solutions for issues are handled consistently with definitions, rules, etc., and are not resolved in a way that could harm data quality. Review issues for opportunities to resolve root cause (preferably in the current build cycle).	Identify root causes of data quality issues and recommend solutions. Record outcomes of issues, especially if resolutions result in residual impact to data quality or need to be included in training or communications.
Perform tests.	Align data elements in user interfaces and reports with glossary definitions and help content. Ensure that usability testing verifies the features supporting data quality are working as designed and cannot be bypassed. Support testers in researching questions about meaning and correct use of data.	Profile and check data prior to, during, and after test loads. Log any deltas between test results and specifications, including definitions and rules. Create feedback loop between teams with dependencies on the data engaged in various types of testing. Help analyze test results and provide feedback on solutions to issues found during testing.

©2013 Danette McGilvray and Masha Bykin

Table 7 — SDLC Phase: Deploy

Deploy Phase Team Activity	Data Governance and Stewardship Activities	Data Quality and Readiness Activities
Move the solution to the production environment and stabilize.	Help to communicate about and resolve data issues found during deployment. Update data artifacts and documentation to reflect changes in the release.	Participate per deployment plans. Identify and research issues found during deployment and stabilization. Help to implement ongoing monitoring of data quality, making use of data quality assessments and testing done during the project.

©2013 Danette McGilvray and Masha Bykin

Table 8 — SDLC Phase: Production Support

Production Support Phase Team Activity	Data Governance and Stewardship Activities	Data Quality and Readiness Activities
Monitor system health, support users, and maintain and enhance iteratively.	<p>Prioritize and make decisions about data quality checks as needed. Ensure accountability for taking action on results.</p> <p>Repeat applicable SDLC activities from Tables 2-7 to maintain and enhance data quality.</p>	<p>Continue to find and recommend additional automated data quality checks as needed.</p> <p>Perform routine data quality assessments.</p> <p>Research, perform root-cause analysis, and propose solutions for data quality issues identified.</p> <p>Repeat applicable SDLC activities from Tables 2-7 to maintain and enhance data quality.</p>

©2013 Danette McGilvray and Masha Bykin

Though the SDLC phases in the tables look like a traditional “waterfall” approach where all steps, activities, and tasks are done in a single sequence, most data quality and governance activities are iterative in nature. To accomplish data readiness within a given project, data-related work efforts often need to go through their individual lifecycles ahead of primary feature builds, and some in parallel. We believe this need for continual iteration is one of the reasons that addressing data quality in waterfall or sequential projects is a challenge. More discussion about iteration will take place after the tables. For now, keep in mind that parallel activities can go on within a single project. When exposing data quality issues, there is often a need to spin off data readiness activities. Later in this report, we will provide an example of how to map these activities to a different SDLC, using an agile Scrum lifecycle.

If your first inclination upon reviewing the activities is “We don’t have time!” please review the myths earlier in the report. Remember, including these activities effectively throughout the SDLC and applying these disciplines as part of all project work will help you keep to your timeline, increase the chances of project success, and lower the risk of business interruption after go-live.

Iteration for Data Readiness

Those used to thinking and tracking projects in a very sequential style may find it unsettling to start data quality and governance activities early in the project, even when all requirements are not known, and then continually make adjustments as more information is learned. However, it is a very effective tactic — if planned and executed well. You can begin data quality assessments as soon as a project is initiated based on knowledge of the business and early project goals. As you discover issues, you can handle them as mini-projects with their

own SDLCs, which you must coordinate with the original project, as shown in Figure 1. The figure illustrates a scenario in which data issues and dependencies are discovered early and resolved prior to related functional development, resulting in a smooth flow of work and avoiding disruptions to plans. This gives time to respond to the myriad data quality issues that will arise, some of which will be very easy to fix if found several months before deployment. Unfortunately, it is more common that these issues exist, but are found later. The worst case is when you discover data quality issues late in functional testing, which poses a risk to the quality of the release and the project timeline. In some cases, these turn into serious emergencies given the short length of time before the move to production.

For example, ACME’s ERP project identified “product” as a needed subject area early in the initiation phase. The project data team worked with the manufacturing business units to start “good housekeeping” data readiness efforts based on known business needs and data quality issues. They flagged unused data such as obsolete items, which would not be migrated to the new system, thereby reducing data volumes and the time it took to extract and test data later in the project. They also started correcting known accuracy issues in attributes, such as unit of measure or item descriptions, so they reflected the actual product. Guidance or requests from project teams helped prioritize these types of issues, so many were taken care of by the time the project team needed the data. They continued data readiness in the early to mid-phases of the project based on actual source data analysis and new requirements. This lessened the number of issues found during testing, but when they arose the team had time to address them quickly. Improvements and issue resolution continued, in some cases up until go-live.

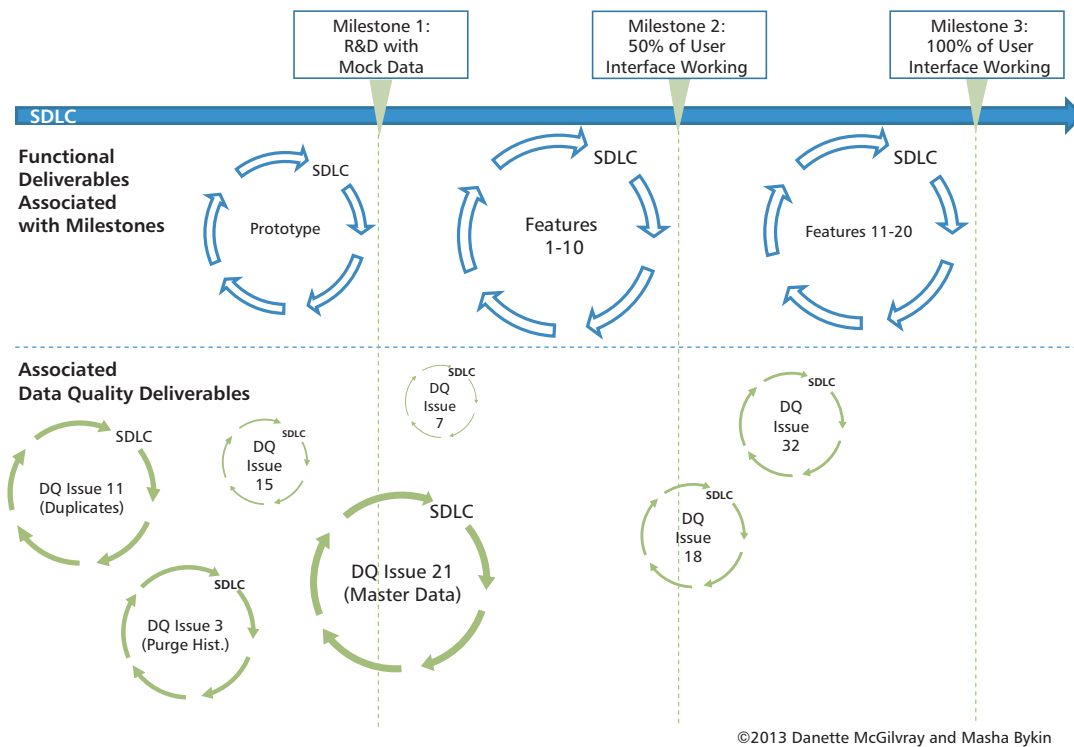


Figure 1 — Data readiness requires iterative and parallel efforts.

The same iterative approach applies to the interaction between those responsible for data governance, stewardship, quality, business processes, and technology. These people will work closely together to research, prioritize issues, and convince others to take necessary action. All need to be aware that they will be involved in a continuous improvement cycle of assessment (e.g., looking at the data and comparing to requirements and goals), awareness (e.g., understanding the true state of the data; identifying gaps, impacts, and root causes), and action (e.g., prevention of future information and data quality problems in addition to correcting current data errors).¹⁰ The iterative nature of data work lends itself well to agile methodologies, which we discuss in the next section.

AGILE WITH DATA GOVERNANCE AND QUALITY

Agile methodologies, like Scrum and Kanban, are highly iterative. Organizations that implement agile practices get good at breaking work down into small units called “stories,” and ordering priorities and dependencies using roadmaps. Roadmap creation and release planning establish high-level priorities and milestones with a focus on tangible outputs and business value. The stories create a context for detailed work plans.

The process of “grooming” is used to make stories small and well defined, and includes sequencing by priority, dependency, and value. These practices involve all the activities that take place in the planning and requirements phases of a waterfall SDLC, as well as some design activities. All the same traditional requirements may be captured at a high level in stories, which are collectively called “a backlog.” All stories are not addressed, however, at the same level of detail or at the same time. The cadence of work is to complete sets of stories in one- to four-week cycles, called “sprints.” Only the stories coming close to inclusion in a sprint get fully defined and detailed. Once selected for a sprint, stories get the full attention of the implementation team, and are further decomposed into tasks. The rest of the backlog continues to be groomed iteratively, with many stories remaining in early stages of definition. In this way, there is heavy attention on small sets of features, and there are frequent small releases of the highest-priority functionality.

With more sequential projects we try to define all requirements before moving on to design. In agile, we create a roadmap and do just enough analysis to expose dependencies and to identify work we can complete in the next sprint or two, knowing that much will be learned during the build. We keep grooming the backlog for future sprints and then update roadmaps

and release plans while working with new information. This gives us the ability to iteratively do stewardship and data quality work in advance of starting the development.

Design starts during story grooming and is a natural place for small and focused data quality assessments and inclusion of data definition work. The discovery of issues during grooming allows for the insertion of new stories into backlogs for issue resolution, and the sequencing of dependent stories appropriately. If data quality assessments are needed, or artifacts need to be produced for large subject areas, these can themselves be stories in the backlog, provided the value is clearly articulated and supported by users and stakeholders.

However we achieve it, when a story relies on data, we need a thorough understanding of the real data and business rules to start and complete that story in one sprint. If that exploration only starts during the sprint, teams will find issues they didn't expect, and will either build workarounds or will be unable to complete the story. These outcomes will reflect negatively on the team, may result in technical debt,¹¹ or lead to issues after releases.

When effective data governance and quality processes are paired with agile methodologies, data definitions and quality measures can help produce clear and complete user stories. Definitions help clarify the information required for the story, while data quality analysis and measures become part of conditions of satisfaction (i.e., how stakeholders will decide the product is good) and can help define test plans. Story readiness criteria (i.e., whether the story is ready for a sprint) can require data elements to have published definitions, to be approved by assigned stewards and available in a company glossary, and to have a readiness assessment completed on source and target data. For example, a story for a screen that creates a new "customer address" would have to reference a definition that explains what is meant by "customer" and whether this is a billing address, shipping address, or a new and different type of address. Some analysis might be required to determine whether the data will be local or international, and whether the data structure accepts the necessary address type.

The iterative nature of backlog grooming fits the iterative nature of definition work and data analysis procedures, which project teams can use to understand and correlate definitions as well as source and target data readiness, mapping rules, and quality tests. When invoking these processes on large quantities of

requirements, they can take a significant amount of time. However, if performed routinely in the course of small and iterative change cycles, much less time is spent and the benefits are immediate and long-lasting. If we establish a strong practice of using the data knowledge network as we write stories and acceptance criteria, our products will iteratively improve data quality, which will then pave the way for greater velocity in both grooming and sprint cycles.

Table 9 provides a mapping of typical agile Scrum activities¹² to related SDLC phases. You can then reference the data governance, stewardship, and quality activities in Tables 2-8 to see where they can fit into an agile Scrum methodology.

When effective data governance and quality processes are paired with agile methodologies, data definitions and quality measures can help produce clear and complete user stories.

BEST PRACTICES FOR COLLABORATION

As a senior manager, you set priorities and determine whether these recommendations are implemented. Ensure all parties engage and coordinate to get data quality the attention it needs at the right time and place in the project. In this section, we highlight some best practices for collaboration.

Data Knowledge Network Collaborating with Project Management

To begin, you must understand how your data knowledge network affects project management:

- Research the project — goals, organization, scope, key players, methodology — to determine how to fit data quality processes, tools, and activities into the project approach.
- Ensure there are people on the project team who are responsible for data deliverables and liaison with the data knowledge network. They must attend the relevant meetings, receive the appropriate communications, and be treated the same as other project team members. Ensure data deliverables are as visible as other project deliverables.
- If your project has separate tracks or work streams (e.g., business processes, technology), we recommend

Table 9 — Agile and Data Knowledge Network Activities in the SDLC

Agile Scrum Activity and Highlights	Same Activities as in Related SDLC Phase Tables 2-8
<i>Envisioning & Product Roadmap</i>	
Focus on product vision and high-value milestones, not end-to-end execution. Identify key roles, skills, resources; form small teams.	Initiation (Table 2) Planning (Table 3)
<i>Release Planning</i>	
Highest-priority stories (based on roadmap) are selected from product backlog for upcoming releases. Focus on one or two upcoming releases, not whole roadmap.	Planning (Table 3) Requirements and analysis (Table 4) Design (Table 5)
<i>Product Backlog Grooming</i>	
Done collaboratively by stakeholders and technical team. Starts during release planning and continues through sprint planning and sprint execution (for upcoming sprints). Stories are written, prioritized, sequenced, and sized into small increments. Iterative clarification of value, design approach, and acceptance criteria until story is ready for sprint.	Requirements and analysis (Table 4) Design (Table 5)
<i>Sprint Planning</i>	
Small number of “ready” stories are sized in more detail and selected as sprint backlog. Sprint backlog decomposed into tasks.	Planning (Table 3) Requirements and analysis (Table 4) Design (Table 5)
<i>Sprint Execution</i>	
Stories in sprint backlog are developed and packaged for release. Testing is part of construction, ideally integrated into build. Implementation exposes new requirements and design issues. Work that can’t be completed goes back into product backlog, sometimes as debt. Retrospectives expose better processes.	Requirements and analysis (Table 4) Design (Table 5) Build & Test (Table 6)
<i>Release</i>	
One or more sets of sprint packages are moved to production. Issues and new requirements are managed iteratively through product backlog.	Deploy (Table 7) Production support (Table 8)

©2013 Danette McGilvray and Masha Bykin

a separate data track. This would be similar to the agile feature or component teams, who have a specific focus but must align and coordinate with other teams. A separate data track keeps a consistent project structure, gives equal emphasis to the data, makes it easier to allocate people, allows them to focus on the data, and ensures the data deliverables are not left out at the whim of teams who are not specialists in this area.

Data Knowledge Network Collaborating with Organizational Change Management

Organizational change management refers to helping teams and individuals accept changes triggered by the project. Accepting change is necessary to achieve lasting benefits from the project. Organizational change management encompasses such areas as organizational alignment and business readiness, communication,

training and documentation, stakeholder engagement, and leader readiness. Note these points:

- If the project triggers organizational changes, ensure the inclusion of appropriate data quality and governance roles and responsibilities.
- Incorporate relevant messages in communications that help prepare the organization for changes.
- Incorporate messages into end-user training. For example, if you train users on how to enter data or create records in the new application, add in scenarios of what happens if data is not entered correctly (immediately to the transaction and to downstream systems). Don't stop with the "how"; also include "why" the procedures are necessary and how they impact the business.
- Ensure leaders obtain the right level of awareness. Managers direct their teams' activities and set priorities. If they don't believe in data quality, it is unlikely their teams will be able to make the necessary efforts.

SUCCESSFUL PROJECTS

ACME leadership prioritized high-quality data as a necessary component to achieve its strategies and formalized its data governance, stewardship, and quality functions. Cynthia and Patel educated themselves in data quality and governance best practices, ensuring their project activities were adapted to leverage the network of resources. Let's take a look at how these changes played out.

Enterprise Resource Planning

The SDLC that ACME used for the ERP implementation included many tasks related to data. However, it was missing several data governance, stewardship, and quality activities. In spite of the fact that the project had already started, the executive leadership team determined that including the data knowledge network in the project would be critical to its success. Cynthia helped raise awareness of how the project would use these resources and who they were. Initially, there was resistance because of concerns these activities would extend the timeline. As the project progressed, however, it was shown that issues were addressed quickly using the data knowledge network and better decisions were being made. Team members soon found that using the data actually saved them time that they could then direct toward other project responsibilities. Cycles of continuous improvement were carried out through all phases of the SDLC.

The project team used a robust third-party data profiling tool to get a view of the actual state of data quality early in the project. As just one example, the project team used the technique of data profiling to get the first broad view of open sales orders in one of the key source systems to be migrated. New system requirements were not identified, yet it was not too early to look at the data based on known business parameters. The company had a policy to have no open sales orders older than six months, and it was assumed this policy was being followed. To everyone's surprise, the data profiling revealed open sales orders dating back several years.

Data stewards worked with business SMEs to understand the impact of the issue. Financial exposure was in the millions of dollars. Investigation found report parameters in the business centers were not set to catch old orders, so there was no visibility to the old open sales orders. Additional root-cause analysis revealed that as data moved between systems, data was corrupted. Solutions included manually closing orders and moving to the history database. Report parameters were corrected to ensure that managers and customer service reps viewed the same information. ACME realized these efforts were not just about data cleanup; rather, it was about addressing the financial impact of incurring the cost of manufacturing and selling its products but not being paid.

Team members soon found that using the data actually saved them time that they could then direct toward other project responsibilities.

This was only one of many examples of continuing attention to data quality that, along with executive leadership providing resources to address the issues, helped the project keep to its timeline. Management realized many of the unpleasant data surprises, if found late in the project during testing, would have made keeping to the timeline impossible. Including data quality activities, using governance and stewardship to ensure the right people were making decisions, and taking action on the data helped prevent disruption to the business once the ERP moved to production. In one case, finance had a clean financial month-end close two weeks after go-live and a clean quarter-end close four weeks later. The project team had heard of other ERP implementations where companies didn't get a clean financial close for several months.

Based on the contributions of the data knowledge network, the company is updating the standard ACME SDLC to include the activities that helped make their ERP implementation a success. Top management has learned the importance of resourcing those activities and the value they provided to the company. Smoother transition to production, shorter stabilization times, and little disruption to business continuity in this project were additional reasons the executive leadership team gave this project their seal of approval.

With a more holistic approach that includes data, project teams will see how they play an important part in ensuring high-quality data and, equally, how high-quality data helps them better fulfill their roles.

Online Experience

Patel observed the formation of the data knowledge network for the ERP project and noticed the efficiencies those activities brought to the team's handling of data issues. He had three Scrum teams using agile methodologies to deliver improvements to customer services in online and internal applications. He organized workshops with the teams to investigate how they were handling data requirements and issues in order to evaluate the need to tap into these resources.

Patel discovered that each team had product backlog items related to customer data changes, which were either not getting good traction in research and design or had stalled for various reasons. Also, some data changes implemented in recent sprints had generated unexpected defects in integrations between applications and exacerbated conflicts in downstream reporting. Patel decided to allocate dedicated time from a project data steward to work closely with the Scrum teams as a consultant. This team member shared best practices for data definitions and modeling, participated in backlog grooming, and reached out to the data stewards and data quality analysis resources as needed.

Data quality analysis exposed integration issues, and the teams concluded there was a strong need for an identity resolution component to better master the customer data. They agreed on a high-level design and put milestones into their roadmaps for the introduction of those capabilities. They changed stories and backlogs to reflect the new design, and all started to build toward

this capability in parallel. Over the course of several sprints, with the help of the stewards and the common glossary of data definitions that had been formed for the ERP, they noticed that data questions were no longer stalling stories.

As stories made their way into sprints, fewer issues were being discovered during testing and in production. Communications about data were flowing between teams and application users, and getting reflected clearly in online help content. Confidence improved in the application changes, the team, and the data. Early analysis of online service improvements and customer feedback showed positive results, encouraging Patel that his commitment to data quality would help ACME deliver better online service to its customers.

CONCLUSION

Over time, the data quality and governance activities described in this report will create “good” information: We can *find* what we need; it is available *when* we need it; it includes *everything* we need; we *understand* it (we don't misinterpret it); and it is *right* (an accurate reflection of the real world). This means we *trust* it and we *use* it ... to make informed decisions and take effective action.

Once project managers, resource managers, stakeholders, and sponsors are aware of how data quality and governance activities can increase project success and decrease the risk of business disruption, those activities will be prioritized, funded, and incorporated into projects. Because the myths have been exposed as false, management can confidently address the data realities. With a more holistic approach that includes data, project teams will see how they play an important part in ensuring high-quality data and, equally, how high-quality data helps them better fulfill their roles. For example, developers will see how their transformation code gets better by knowing the actual state of the data, business analysts will find ways to contribute to these improvements, and testing teams can plan their work with confidence knowing they won't be the first to discover all data quality issues. All roles will make use of the data knowledge network to analyze, troubleshoot, identify root causes, prioritize, and implement solutions.

Of course, the real benefits come, not from being able to say we have high-quality, integrated information, but the fact that the business can make informed decisions and take effective action. This is why we never invest in data quality for the sake of data quality. Focus your

quality improvements efforts on the data most important to your business. Projects are a good way to do that because they reflect organizational priorities. If you can incorporate data quality into projects, you will not only make projects more successful, but improve the quality of the data that matters most.

Have you activated your data knowledge network? Have you included data quality and governance into your projects? What might you do better? What are your next steps? How might your organization benefit? Share your experiences and tell us what you think. You can reach Danette McGilvray at danette@gfalls.com and Masha Bykin on LinkedIn.

ENDNOTES

¹While we use the words “business” or “company” in this report, everything said here applies to any type of organization — for profits, nonprofits, government, education, and health-care — because all invest in projects and all depend on information to successfully provide their products or services.

²We see data as known facts or other items of interest to the business, and information as those facts in context. Other than these distinctions, this report does not generally differentiate between data and information. Some organizations respond to “data quality,” others to “information quality.” Use the most meaningful term based on your audience.

³“Software Testing, Economics” (Wikipedia): “It is commonly believed that the earlier a defect is found, the cheaper it is to fix it. ... For example, if a problem in the requirements is found only post-release, then it would cost 10-100 times more to fix than if it had already been found by the requirements review.”

⁴McGilvray, Danette. *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information*. Morgan Kaufmann, 2008, p. 52.

⁵Adapted from Early, Susan (ed.). *The DAMA Dictionary of Data Management*. 2nd edition. Technics Publications, 2011, p. 88.

⁶McGilvray (see 4), p. 5.

⁷McGilvray (see 4), pp. 30-33.

⁸Olson, Jack E. *Data Quality: The Accuracy Dimension*. Morgan Kaufmann, 2003, pp. 13-16.

⁹McGilvray (see 4), pp. 246-251.

¹⁰McGilvray (see 4), p. 55.

¹¹“Technical Debt.” (Wikipedia): “Technical debt (also known as design debt or code debt) is a metaphor referring to the eventual consequences of poor or evolving software architecture and software development within a codebase. The debt can be thought of as work that needs to be done before a particular job can be considered complete. As a change is started, there is often the need to make other coordinated changes at the same time in other parts of the codebase or documentation. The other required, but uncompleted changes, are considered debt that must be paid at some point in the future.” Note that Cutter

Consortium offers a Technical Debt Assessment and Valuation engagement (www.cutter.com/consulting-and-training/technical-debt-assessment.html).

¹²Rubin, Kenneth S. *Essential Scrum: A Practical Guide to the Most Popular Agile Process*. Addison-Wesley Professional, 2012.

ABOUT THE AUTHORS

Danette McGilvray is President and Principal of Granite Falls Consulting, Inc., a firm that helps organizations increase their success by addressing the information quality and data governance aspects of their business efforts. Focusing on bottom-line results, Ms. McGilvray and her team promote the intentional management of information assets with the same rigor used to manage other company assets such as products and human and financial resources. A skilled facilitator, program, and project manager, she has worked with people in all levels of the organization and from all functional areas, giving her a valuable perspective on organizational challenges. Ms. McGilvray also emphasizes communication and the human aspect of information quality and governance.

Ms. McGilvray is the author of *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information*. An internationally respected expert, her Ten Steps™ approach to information quality has been embraced as a proven method for creating and managing information and data quality projects in the enterprise. This approach, in which she has consulted with *Fortune* 500 clients and trained thousands of workshop attendees, applies to all types of data and all organizations. It is the first data quality book translated into Chinese. Ms. McGilvray’s chapter “Data Quality Projects and Programs” in the *Handbook of Data Quality* explains the fundamentals of successful data quality projects and formal data quality programs. She provides practical solutions based on real situations and challenges. In addition, the International Association for Information and Data Quality (IAIDQ) awarded Ms. McGilvray with its Distinguished Member Award in recognition of her “outstanding contributions to the field of information and data quality.” She is also the recipient of the Utah State University Jon M. Huntsman School of Business Professional Achievement Award. Ms. McGilvray has been profiled in *PC Week* and *HP Measure Magazine*. She is an invited speaker around the world and has visited China to discuss roles and opportunities for women in the computer field. She can be reached at danette@gfalls.com.

Masha Bykin has spent two decades practicing software engineering, IT architecture, data warehousing, and business intelligence in the diverse industries of contact center, software, automotive, and insurance services. She spent three years on a data governance program, working directly with business to initiate and evolve the program, which supported many projects and had broad organizational participation. In her current role as Data Delivery Manager at AAA of Northern CA, NV, and UT, Ms. Bykin leads the development of company-wide data services. She holds a Certified Data Management Professional (CDMP) certification at the mastery level and is working on an information science graduate degree from San Jose State University. She can be reached via LinkedIn.



Intense, Interactive Instruction. Effective Learning.

Register today and save!

Single Seat: \$1995 for a limited time (save \$500!)

Team Builder: Buy 1 seat for \$2495 and bring a colleague at a deep discount. (Save 30%)

Executive education on IT leadership and emerging trends. A year's worth of professional development and personal enrichment in 3 invigorating days.

From an in-depth case study taught in the popular business school style, to interactive small group exercises and keynotes that help you identify business opportunities made possible by emerging technologies, you'll enjoy truly unbiased discussion and meaningful debate on today's IT opportunities and challenges at the Cutter *Summit*. Discover and learn about new strategies, technologies, and leadership skills — from Cutter's exceptional lineup of experts — that will help you embrace the ever-unfolding opportunities and challenges of the SMAC business environment.

In addition, you'll benefit from hands-on seminars and roundtables led by Cutter's Practice Directors and Senior Consultants on topics such as software engineering and agility, business and enterprise architecture, CIO/CTO issues, and data insight and social BI, to name a few.

You'll enjoy (and join in on!) raucous panel debates; networking at lunches, breaks, and entertaining evening events; and get one-on-one guidance and input from expert presenters and participants.

Monday, 4 November 2013

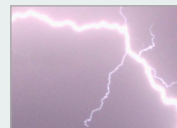


The Evolving Role of 21st-Century Technology Leaders

Keynote by Robert D. Scott

Cutter Fellow; Director of the Information Systems Executive Forum, Ross School of Business, University of Michigan

Panelists: Robert Austin



Lightning Talks

Hosted by Tim Lister

Short and to-the-point presentations around a single strategy, technique, or success story.



Evening Cocktail Party

Unwind and socialize with the speakers and your fellow attendees while enjoying some of Boston's tastiest regional specialties.



Big — and Fast — Data Analytics

Case Study with Vince Kellen

Cutter Fellow; Senior Vice Provost for Academic Planning, Analytics & Technologies, University of Kentucky

Tuesday, 5 November 2013

**Sustainable Growth: Achieve It with Highly Motivated Teams**

Roundtable with Lynne Ellyn

Cutter Fellow**Digging for Gold in the Emerging Technology Pile of Hype**

Keynote by Lou Mazzucchelli

Cutter Fellow**Agile in the API Economy**

Keynote by Israel Gat

Fellow and Director, Cutter Agile Product & Project Management Practice**Putting Your Leadership Skills to the Test**

Active Learning Exercises with Michael Roberto

Cutter Fellow; Trustee Professor of Management at Bryant University

Wednesday, 6 November 2013

**The Chief Data Officer**

Roundtable by Larissa Moss

Senior Consultant, Cutter Consortium

Choose a Track ►

BUSINESS TECHNOLOGY STRATEGIES TRACK**CIO/CTO Roundtable**

Forum with Ron Blitstein

*Fellow and Director, Cutter Business Technology Strategies Practice***AGILE TRACK****Agile, Software Engineering, and Product Development**

Workshop

*TBA***DATA INSIGHT & SOCIAL BI TRACK****Data Insight and Social BI**

Workshop

*TBA***BUSINESS & ENTERPRISE ARCHITECTURE TRACK****Business and Enterprise Architecture**

Workshop with Mike Rosen

Director, Cutter Business & Enterprise Architecture Practice**A Theory of Practice: Soft Decision-Making in the Context of a High-Pressure IT Organization**

Keynote by Tom DeMarco

Cutter Fellow●●● See the full program at
www.cutter.com/summit.html

Data Insight & Social BI

Decision makers need more support than ever — and they need it now. Analysis of structured corporate operational data is no longer a competitive differentiator. Today's decision makers also need the ability to evaluate market sentiment, identify early indicators of group behavior changes, and respond to events before they make front-page news. Business intelligence as usual is no longer sufficient.

Data visualization, push notifications, exceptions alerts, critical event processing ... these tools will help you avoid drowning in the ocean of Big Data — structured/unstructured, internal/external — and meet the increasing demands of decision makers. From collaboration to crowdsourcing, Agile data warehousing to master data management, analytics, mobile computing, knowledge management, and social media, Cutter's Data Insight & Social BI experts provide unparalleled guidance through original research, consulting, training, and executive education.

Cutter's Data Insight & Social BI team comprises thought leaders and pioneers in the diverse business intelligence arena. These experts have developed methodologies that are revolutionizing data management and being implemented in organizations worldwide. Our team has written acclaimed books, lectured, coached, and trained thousands of professionals who have benefited from their seasoned expertise in data warehousing, Agile BI, knowledge management, collaboration, social networking and media, crowdsourcing, and more.

You'll find extensive forums for interacting with Cutter's experts to brainstorm, debate, and learn. From research and inquiry privileges to regular virtual meetings with Cutter's Data Insight & Social BI thought leaders, participation in webinars, and Q&A sessions, Cutter will help you discover new strategies and gain insight, enabling you to maximize your most valuable assets: data and knowledge.

Products and Services Available from the Data Insight & Social BI Practice

- The Data Insight & Social BI Membership
- Research & Analysis
- Inquiry Response
- Consulting & Mentoring
- Inhouse Training & Executive Education
- Research Reports

Cutter Consortium Practices

Each Cutter Consortium practice includes a subscription-based research service, plus consulting, training, and executive education services:

- Agile Product & Project Management
- Business & Enterprise Architecture
- Business Technology Strategies
- Data Insight & Social BI

Senior Consultant Team

The Cutter Consortium Data Insight & Social BI Senior Consultant Team includes the trailblazers in Agile data warehousing/ business intelligence, as well as experts in the data analytics, knowledge management, social networking, and collaboration fields. You get sound advice and cutting-edge tips, as well as case studies and data analysis from best-in-class experts. This brain trust includes:

- Verna Allee
- Claude Baudoin
- David Coleman
- Lance Dublin
- Joseph Feller
- Curt Hall
- Dave Higgins
- Ralph Hughes
- Larissa T. Moss
- Ciaran Murphy
- San Murugesan
- Ken Orr
- Thomas C. Redman
- Annie Shum
- Karl M. Wiig