Executive Summary

Data Quality and Governance in Projects:
Knowledge in Action

by Danette McGilvray and Masha Bykin

Organizations invest substantial resources in projects. From exciting new innovations to mundane, yet necessary, improvements, companies institute projects to help the business address issues, support goals, or take advantage of opportunities. 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. The accompanying *Executive Report* describes how managers can increase the success of their project portfolios by making data quality, governance, and stewardship resources and activities an integral part of a project solution (or software or systems) development lifecycle, or SDLC. Incorporating this focus into projects is a critical component of treating data as a valuable asset, deliberately managing it as we do our other assets like people, real estate, inventory, and money.

In the report, we introduce the data knowledge network as the combination of the data governance, stewardship, and quality resources in an organization. We also discuss the data quality assessment techniques that expose gaps between existing data and new requirements. Finally, we examine the concept of data readiness, showing how it ensures that data within a project scope is prepared to meet the business requirements and needs of the new solution.

Examples from ACME, a fictional manufacturer and distributor of health products, illustrate real challenges that companies face in their projects. In addition, the report discusses why businesspeople often overlook data as well as seven myths about data in projects. Believing such myths causes suboptimal project results at best, and failure at worst, because reality is not accounted for in plans and budget.

WHY WE NEED THE DATA KNOWLEDGE NETWORK

Many projects concentrate their efforts on implementing technology used to support the business needs (replace System X with System Y) or deliver new functionality. In meetings, the topic of conversation is usually about which application to purchase and which vendor to choose. 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.

Lack of knowledge of the interdependencies of the data between systems results in fixing some data, while breaking data elsewhere. Unpleasant surprises come late in the project during testing; for example, when data quality issues are uncovered. Addressing the issues at this point pushes out the project timeline and puts a strain on financial and human resources. The cost is even higher when a manufacturing plant cannot get products out the door after the new ERP goes live, high visibility compliance reports are incorrect, or customers turn to social media to air their complaints about issues with online services. These are examples of problems that plague projects that neglect to address the data and information aspects of their efforts.

The good news is, your organization can do better, you already have the ingredients, and what you learn from the report will help. It includes reference tables that describe critical activities related to data quality and governance in each phase of a typical SDLC — initiation, planning, requirements and analysis, design, build and test, deploy, and production support.

Though some SDLCs are sequential, such as the "waterfall" approach, most data quality and

governance activities are iterative in nature. We believe this need for continual iteration is one of the reasons that addressing data quality in waterfall or sequential projects is a challenge. Those used to thinking of projects in a sequential style often find it unsettling to start data quality activities early in a project when all requirements are not known. Yet, you can begin data corrections early in the project based on current business needs. For example, as soon as you define "product" as a needed subject area, you can address known issues, such as corrections to unit of measure or item descriptions, so they reflect the actual product. This is good housekeeping.

Whenever data quality issues are exposed, there is often a need to spin off data readiness activities. These are like mini-projects with their own SDLCs, which need to be coordinated with the original project. You can then continue data readiness in the mid-phases of the project based on actual source data analysis and new requirements. Additional changes can be made based on test results, and can continue until go-live. This type of iteration is a very effective tactic — if planned and executed well.

The same iterative approach applies to 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).[1]

The iterative nature of data work lends itself well to agile methodologies. The report also summarizes the aspects of agile where data quality and governance activities naturally fit. To further clarify how those activities fit into an agile Scrum methodology, a table maps the agile Scrum activities to the related SDLC phases. We wrap up the report with specific examples of our case study ACME, which illustrate the benefits of including the data knowledge network in projects.

CONCLUSION

You *will* find data quality issues that adversely impact the project and must be addressed — it's just a matter of *when*. Proactively addressing data quality helps you better manage your timeline, people, and money. As a senior manager, you set priorities. Your commitment will determine whether data quality gets the attention it needs at the right time and place in your projects.

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.

ENDNOTE

[1]McGilvray, Danette. *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information*. Morgan Kaufmann, 2008.

Data Insight & Social BI Practice

FOR MORE INFORMATION

For more information on Cutter Consortium's Data Insight & Social BI practice and other services, contact: Tel: +1 781 648 8700; Fax: +1 781 648 8707; Email: sales@cutter.com.