White Paper

A Capabilities-Based Approach to Data Quality

By Robert Grant Beauchamp

Contents

White Paper

A Capability-Based Approach to Data Quality

By Robert Grant Beauchamp

Introduction

The collective corporate SharePoint landscape is littered with the detritus of past data quality initiatives. A data dictionary here, an out-of-date schema there, along with an incomplete list of data stewards. If the organization is lucky, it has a pretty good warehouse containing its master data. Although many who might take advantage of it are vague on its contents.

It is difficult to implement a data quality initiative and even harder to sustain its good work over time. Effective management and governance of data requires the development of new policies, processes, procedures, roles and responsibilities. Implementing them requires laying groundwork, educating stakeholders and building the political support necessary to attain budge approval. A well implemented change management program may be necessary ensure full adoption.

As organizations enter the world of artificial intelligence, machine learning, the internet of things and advanced analytics the quality of their data and the ability to leverage it is more important than ever.

Therefore, it's critical that additional ways are found to improve the quality of an organization's data. One alternative is to attack the data quality problem from a different angle and use a capability-based approach.

Purpose

The purpose of this document is to provide a:

- Definition of a data-related activity and capability.
- Method to identify, classify and assess data-related activities and capabilities.
- Method for implementing plans to improve capabilities
- Rationale for taking a capabilities-based approach to data quality.

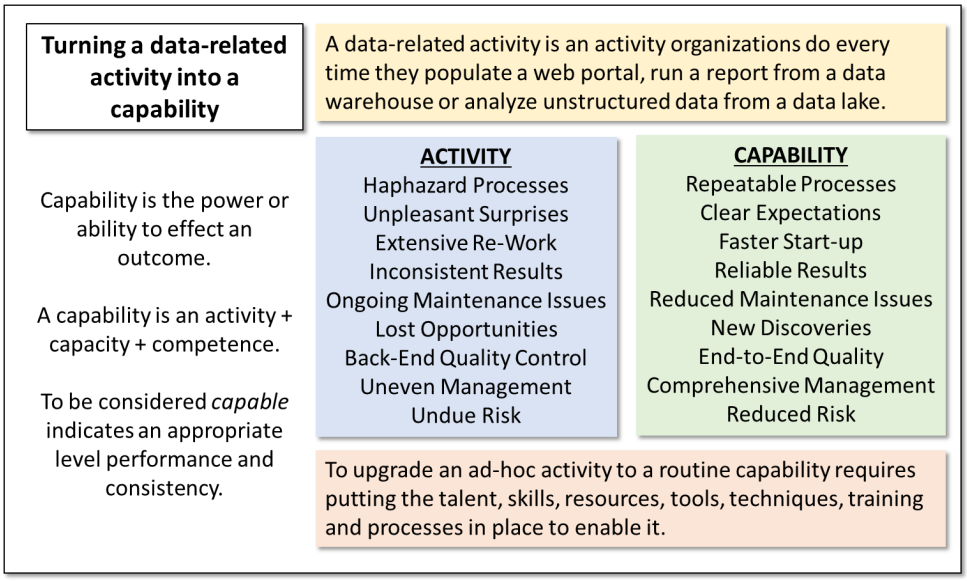
Activities vs. Capabilities

The capability-based approach starts with a simple premise:

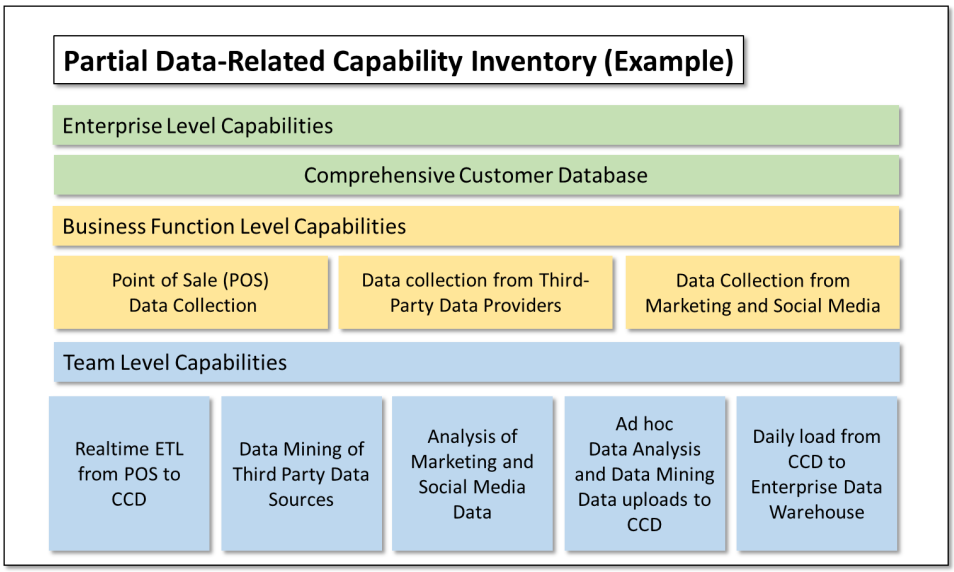
Improve the performance of an organization's data-related activities and the quality of the organization's data will improve.

Data related activities include anything that enters, modifies, moves, retrieves, shares or stores data. Data related activities include everything from data entry and data sourcing to data warehousing, ETL and advanced analytics. It is what the organization is doing every hour of every day that impacts data in one way or another.

However, just because an organization is engaged in an activity, does not mean the activity is being performed well. There is an important distinction between a data-related activity and a data-related capability. One definition of the word capability is 'the power or ability to effect an outcome.' If an activity is not consistently generating the desired outcome, it is just that, an activity, it is not a capability.



Activities vs. Capabilities (Continued)	A good rule of thumb is that ***a capability is an activity plus capacity plus competence***. Capacity represents the ***amount*** of resources required and competence represents the ***level*** of skill required. Remember, to be considered ***capable*** indicates an appropriate level of performance and consistency. Knowing how to distinguish between an organization's data-related activities and its data-related capabilities is an important early step towards data quality improvement.
The Data-Related Activity and Capability Inventory	Another important early step is to start identifying and documenting the organization's data-related activities and capabilities. As data-related activities are identified, the list gets very long very quickly. The inventory can be a simple list or multiple lists. It is helpful, however, to create a structured inventory that presents the activities in logical groupings. Activities can be identified at the Enterprise, Business Function or Team levels. Activities at the Enterprise and Business Function levels are usually supported by team level activities. A document showing the supporting relationships between the team level activities and the Business Function and Enterprise-level activities is a very powerful tool. It allows stakeholders see relationships in ways they never could before.



The Data-Related Activity and Capability Inventory (Continued)	It is important to not get distracted trying to scientifically determine if an activity is at the enterprise, business function or team level. Or whether one large activity should be broken into multiple activities. Or even if an activity is a capability. Use the names that the people who perform the activities use. The most important thing early on is to identify and document them as best one can. The complex relationships between activities and capabilities will reveal themselves over time, as more activities and capabilities are 'discovered.' As activities are upgraded to capabilities, the Activity Inventory will begin to look more like a Capability Inventory.
Seeing Activities and Capabilities as Stakeholders See Them	Once the data-related activities inventory is under way, it is important to begin to classify the performance of the activities. Again, it is important early on to not get distracted trying to classify data-related activities with complicated or comprehensive sets of objective standards. Industry standards and industry best practices are not equivalent to what the organization needs and how to best meet those needs. Rather, capabilities are judged on their ability to deliver what is required by the organization in the way that works best for the organization. And the judges are the organization's stakeholders. One does not have to travel too far or probe too deeply to learn how a stakeholder feels about the performance of a data-related activity. Most stakeholders are more than willing to describe, in much detail, how their needs are not being met. This is not the time for defensively pulling out SLAs. This is the time for listening and understanding and keeping things easy. Classifying a data-related activity is the first step in developing a plan for improving that activity and setting it on the road to becoming a capability. Where capabilities do indeed exist, classification by the stakeholders puts them on the path to improvement.

Classifying the performance of activities and capabilities

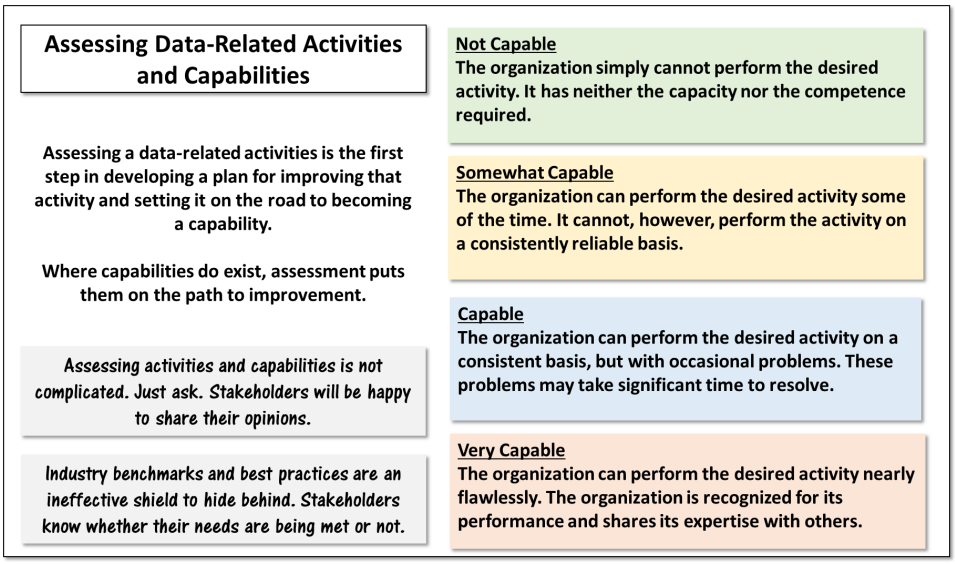
When walking through the activity inventory with a stakeholder, classifications can be as simple as:

Not Capable
The organization simply cannot perform the desired activity. It has neither the capacity nor competence required.

Somewhat Capable.
The organization can perform the desired activity some of the time. It cannot, however, perform the activity on a consistently reliable basis.

Capable
The organization can perform the desired activity on a consistent basis, but with occasional problems. These problems may take significant time to resolve.

Very Capable
The organization can perform the desired activity nearly flawlessly. The organization is recognized for its performance and shares its expertise with others.



Understanding Performance Variables

Whether the quality of data is good or poor, how it got that way usually has little to do with the data itself. Data-related activities, like most business activities, are subject to the vagaries of the day-to-day demands of running of an organization.

People generally do the best they can with the resources available, the time allotted and the direction they receive. Even their best efforts, however, will often fall short when pressures become too great.

The level of performance of a data-related activity is dependent on several variables. A variable is a quantity or quality that changes with context. Variables will impact performance negatively or positively to different degrees. Hence, why they are called variables.

These variables include:

- Immediacy
- Timeframes
- Resources
- Priorities
- Data knowledge
- Tools and technology
- Talent and trained resources
- Processes, procedures and standards.

There are also variables that might be unique to an organization's industry or how the organization itself is run.

Unfortunately, the first four variables represent the realities of running any organization and often can't be avoided. Sometimes they are characteristic of an organization's corporate culture, sometimes of the market. Either way, they are extremely difficult to mitigate directly.

Fortunately, it is much easier to directly address the last four variables. Moving these variables from a negative to a positive is what the capabilities-based approach is all about.

Addressing the Variables

Variables Resistant to Change

It is important to understand the profound impact variables have on the quality of data and the performance of data-related activities. Some variables are not data-specific or technologically exotic but still negatively impact data quality.

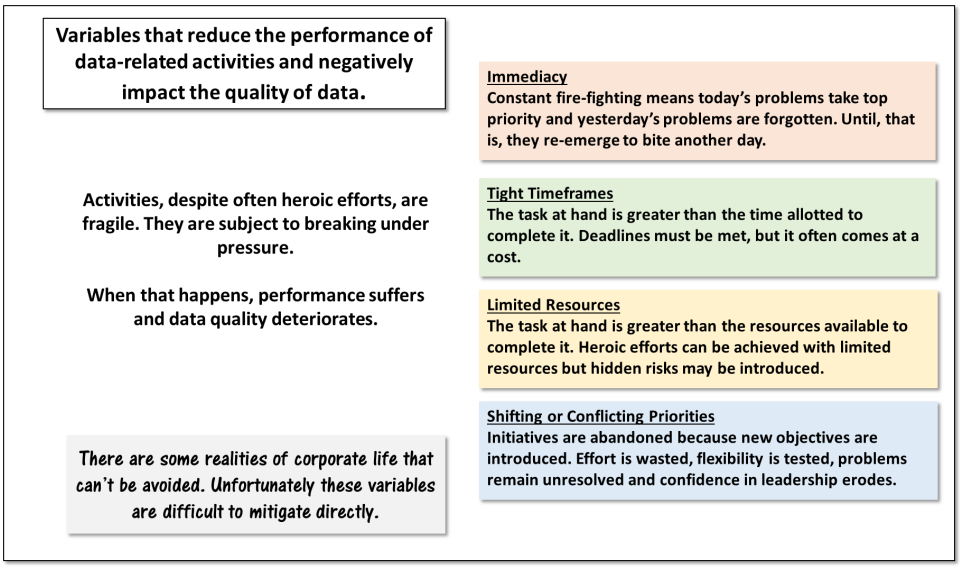
Never-ending Immediacy
Constant fire-fighting means today's immediate problems take top priority and yesterday's problems are forgotten. Until, that is, they re-emerge to bite another day.

Tight Timeframes
The task at hand is greater than the time allotted to complete it. Deadlines must be met, but it often comes at a cost.

Limited Resources
The task at hand is greater than the resources available to complete it. Heroic efforts can be achieved with limited resources, but hidden risks may be introduced.

Shifting or Conflicting Priorities
Initiatives are abandoned because new objectives are introduced. Effort is wasted, flexibility is tested, problems remain unresolved and confidence in leadership erodes.

Any attempt to move these variables from a negative to a positive is an enterprise-level effort. A data quality initiative can certainly contribute, but it is not responsible for, nor should it be distracted by this daunting challenge.



Variables Amenable to Change

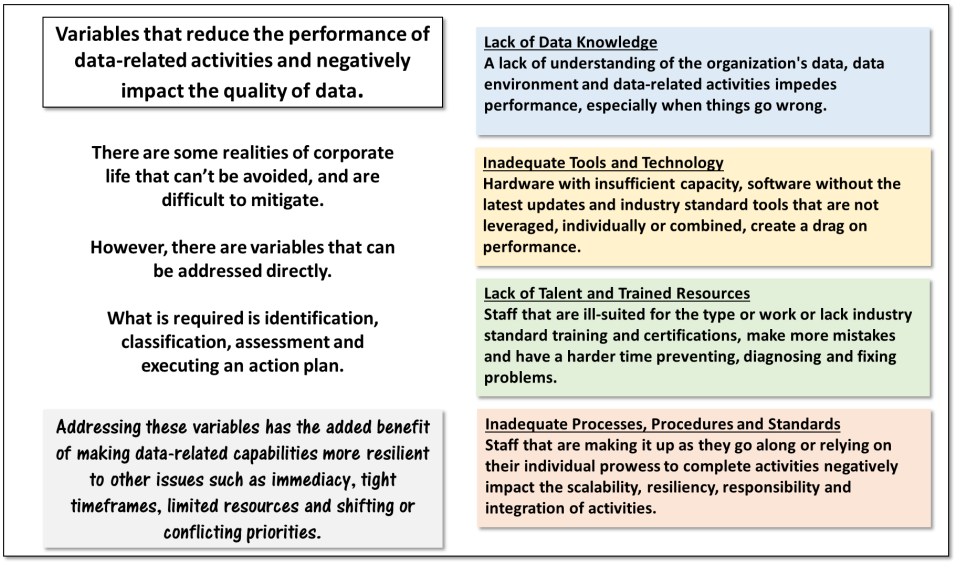
Fortunately, while some issues such as shifting or conflicting priorities can only be indirectly mitigated, it is much easier to address the next four variables directly. Most often, the poor performance of a capability is due to neglect of one of these variables. Types of neglect include:

Lack of Data Knowledge
A lack of understanding of the organization's data, data environment and data-related activities impedes performance, especially when things go wrong.

Inadequate Tools and Technology
Hardware with insufficient capacity, software without the latest updates and industry standard tools that are not leveraged, individually or combined, create a drag on performance.

Lack of Talent and Trained Resources
Staff that are ill-suited for the type or work or lack industry standard training and certifications, make more mistakes and have a harder time preventing, diagnosing and fixing problems.

Underdeveloped Processes, Procedures and Standards
Staff that are making it up as they go along or relying on their individual prowess to complete activities negatively impact the scalability, resiliency, responsibility and integration of activities.



Solving for 'X'

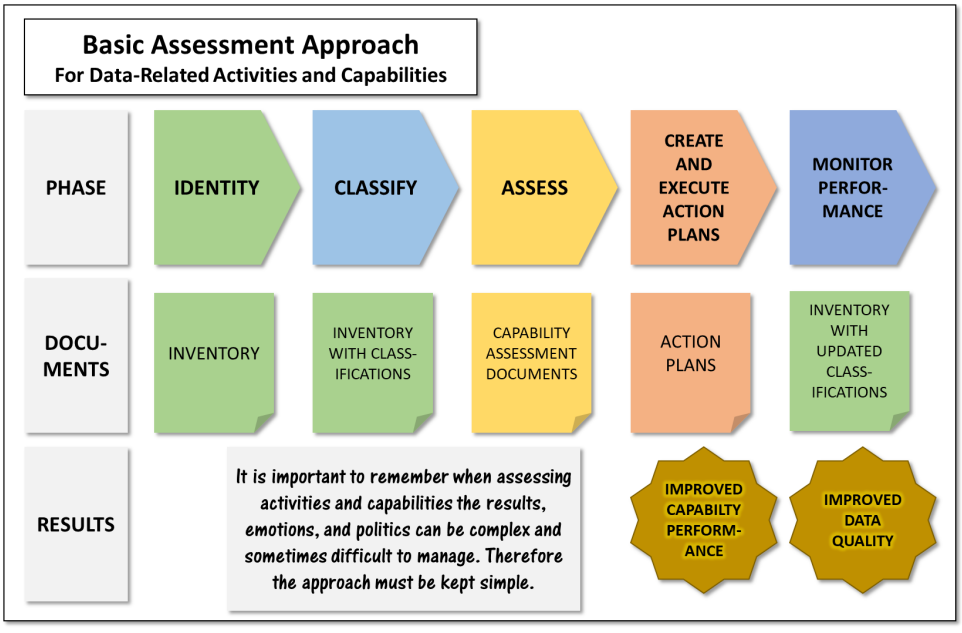
A Simple and Direct Approach to Improve Data-Related Activities and Capabilities

At its core, a capability-based approach asks four simple questions:

- How well is a data-related activity being performed?
- If not performed well how does that impact the quality of the organization's data?
- What needs to be done to improve the performance of the activity to bring it up to the level of a capability?
- If a capability does indeed exist, what needs to be done to improve it?

While the questions may be simple, the answers are often not. The answers may be technically complex. The answers may bring out unpleasant emotions. The answers may uncover issues that have been long ignored or actively covered-up.

Because the answers, emotions, and politics are complex, and can be difficult to manage, the approach for getting those answers must be kept simple.



Listen to those that know

When ***classifying*** capabilities, it is very important to listen to stakeholders. When ***assessing*** capabilities, it is imperative to listen to the people who are performing those activities every day.

There are times when expertise from outside the organization is needed. However, more-often-than-not, the people doing the work every day already have a good list of shortcomings and what needs to be done to fix them. They just need to be listened to and taken seriously.

The Starter Questions

One way to keep things simple, when assessing an activity or capability, is to start with a few basic questions. These questions are divided into four categories.

Data Knowledge

- **Has all the data that is impacted by the capability been identified?**
- **Is the data defined in a data dictionary?**
- **Have all the locations where the data resides, is modified, or sent been identified?**
- **Are there up-to-date schemas and models available?**
- **Are there data flow maps, describing upstream and downstream impacts?**

Tools and Technology

- **Are there tools available that aren't being utilized?**
- **Are there industry standard technologies that need to be adopted?**
- **Does new hardware need to be purchased, new software installed, or a new host location established to improve system performance?**

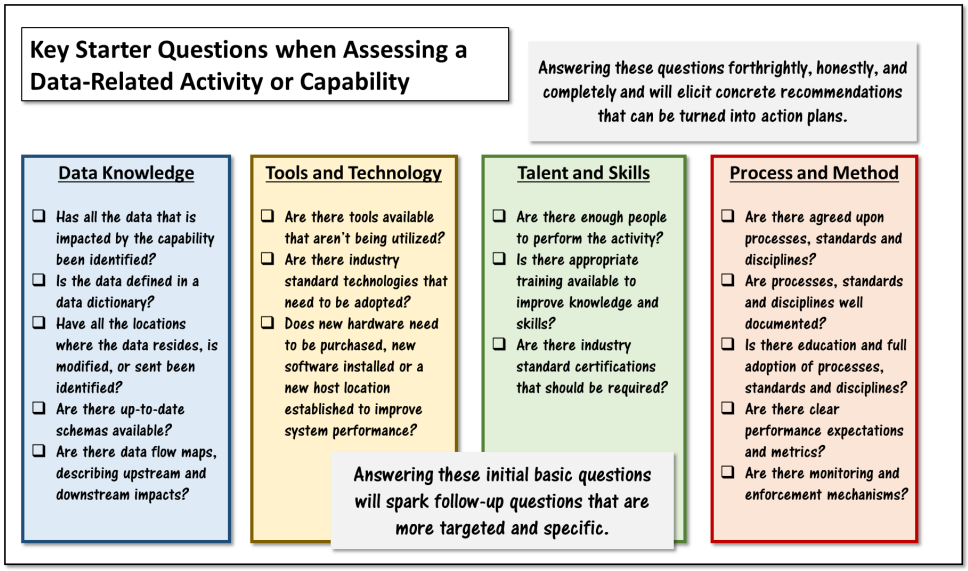
Talent and Skills

- **Are there enough people to perform the activity?**
- **Is there appropriate training available to improve knowledge and skills?**
- **Are there industry standard certifications that should be required?**

Process and Method

- **Are there agreed upon processes, standards and disciplines?**
- **Are processes, standards and disciplines well documented?**
- **Is there education and full adoption of processes, standards and disciplines?**
- **Are there clear performance expectations and metrics?**
- **Are there monitoring and enforcement mechanisms?**

The Starter Questions (Continued)	Answering these questions will spark follow-up questions that are more targeted and specific. Answering these questions forthrightly, honestly, and completely and will elicit concrete recommendations that can be turned into action plans. Action plans turn the variables from negative to positive. Executing the action plans is the solution for x.



A Simple Example of an Assessment

The Case of the Reluctant Night Run Manager

	The Data Quality Improvement Team has decided to take on the data-related capability cited by the stakeholders as the most frustrating to them. The Night Run. The stakeholder classification of this capability is "Somewhat Capable" as it is performed successfully most of the time. The manager of the night run, Jean, takes a lot of heat from stakeholders when the night run fails. Criticism seems to be directed only at her and not at her leadership or other data staff. This is probably because Jean and the Night Run have, over the years, become synonymous in the minds of many. When addressing issues in the past, it has been stated that Jean is resistant to change. Scuttlebutt suggests that she keeps knowledge to herself, to make herself indispensable.
Assessment Results	The Data Quality team has conducted interviews using the starter questions, probed deeper with follow-up questions and have reviewed existing documentation. These are the results of the assessment as documented by the team. **Capability Under Review** Performing the Night Run **Level of Capability** Somewhat Capable **Current State** Night run fails approx. twice a month. **Desired Future State** Night run is successful every time. **Data Knowledge** • The jobs that are part of the night run need to be run in a specific order, so that an ancillary system can get updated, make calculations and return data to the main system where the updated data is sent to the next ancillary system.

- If jobs are run out of order or fail, there is a cascading effect on the accuracy of the data in all the systems. This requires an extensive re-running of jobs, for which there is only time on the weekends. It may take an entire week to recover. During that time data is not up-to-date and is out of sync between systems.

Tools and Technology

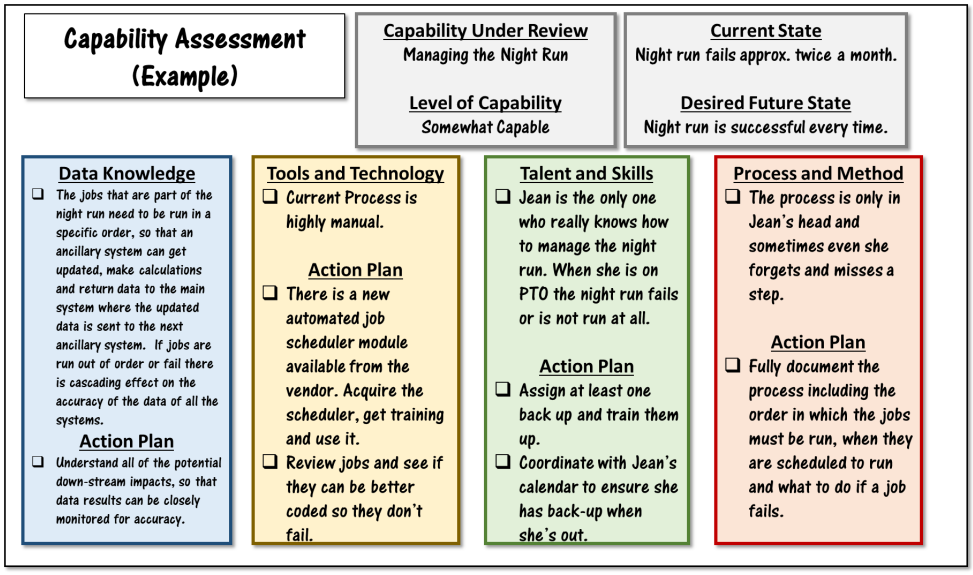
- Current process is highly manual.

Talent and Skills

- Jean is the only one who really knows how to manage the night run. When she is on PTO the night run fails or is not run at all.
- As it turns out, Jean is not resistant to change and has asked for help in the past. If she seems controlling it is because the night run has become much more complex over the years. So complex that when someone else tries to cover for her, it invariably fails, and she takes the blame.

Process and Method

- The process is only in Jean's head and sometimes even she forgets and misses a step.



Action Plans	The Data Quality Team worked with Jean, Jean's leadership and a few others to develop action plans that change the variables. These are the action plans as documented by the team. **Data Knowledge Action Plan** • Understand and document, with a data flow diagram, all the potential down-stream impacts, so that data results can be closely monitored for accuracy. **Tools and Technology Action Plan** • There is a new automated job scheduler module available from the vendor. Acquire the scheduler, get training and use it. • Review jobs and see if they can be better coded so they don't fail. **Talent and Skills Action Plan** • Assign at least one back up and train them up. • Coordinate with Jean's calendar to ensure she has a trained back-up when she's out. **Process and Method Action Plan** • Fully document the process including the order in which the jobs must be run, when they are scheduled to run and what to do if a job fails.
Advantages to the Assessment	There are advantages to taking the time and effort to conduct this type of simple assessment. They include: • Understanding the problem before determining the solution. • Gaining knowledge first-hand to dispel myths, rumors, gossip. • Correcting past faulty reasoning and sub-par analysis. • Avoiding consultant-speak and complex methodologies that are impossible to fully comprehend and implement.

Activities are Fragile, Capabilities are Resilient

The Case for a Capability-based Approach to Data Quality

A capability-based approach to data quality does not negate the need for data management or data governance. These are required. However, getting them up and running is full of organizational challenges that are not easy to overcome.

Unfortunately, too many well-meaning data quality initiatives fall victim to such forces as staff turnover, inconsistent adoption, uneven rigor across business silos, the level of commitment to maintain them and good old-fashioned resistance to change.

Low barriers to entry
Compared to enterprise-level data quality initiatives, the barriers to a capability-based approach are low. Identification, classification, assessment and implementation of action plans can begin with a small team targeting a few well-known problems.

As stakeholders begin to see and talk about the results, their peers will want to sign on. The more capabilities the team tackles, the better the results will be, which in turn creates more interest.

Lower levels of risk
All large enterprise-wide programs carry risk. If a large program is founded on a certain set of assumptions, and those assumptions turn out not to be true, resources are wasted, and reputations are tarnished. By starting small there is the opportunity to discover what works and what doesn't for an organization. It is easier to challenge an assumption when the stakes are smaller. If things get off-track and need to be adjusted, fewer resources have been expended.

Resiliency
Capabilities are resilient, comprising mechanisms that account for, and can absorb, the pressures an organization deals with every day. Capabilities act as a stabilizer. Data knowledge, talent, tools, processes and standards absorb the waves of uncertainty, crises and shifting priorities.

Action plans are targeted and specific
Rather than the perennial 'We don't have enough people,' 'We need to increase our training budget' or 'Why can't I purchase

	and install this non-compliant software?' capability-based action plans provides specific and targeted direction. It becomes much easier to get budget approval when a clear line can be drawn between the need, the solution and the request. **Action plans are modular and flexible** Action plans can be tackled at a pace with which the organization is comfortable. The organization can delegate them individually to the relevant teams, or it can bundle them into larger, more comprehensive projects and programs. **Implementing action plans is iterative** At the end of most large projects or programs, there is an exercise called 'Lessons Learned.' The hope is that in the future the organization will have learned these lessons and avoid the same mistakes in the future. With a capability-based approach, implementation is iterative. Lessons are learned more quickly and can be mitigated more quickly. **Quality begins from the bottom up and from the start.** If the last thirty years of quality theory and practice have proved anything it is that quality cannot be imposed from above or tacked on at the end. It is easier to change behavior than to change attitudes. Quality is about how the job is done. Help data professionals do their jobs better and data quality will improve.

About the Author

Robert Grant Beauchamp is a consultant, architect, and former CIO with a proven record of helping organizations understand and improve their data quality, data environments and data-based capabilities. As a systems integrator, Robert has successfully introduced and implemented data-related technologies such as BI, EDI, ETL, data warehousing and three-tier architectures.

If you would like to learn more about data quality or would like help improving data quality within your organization, connect with him on LinkedIn or at datahust.com.

Robert has filled the roles of computer journalist, tech writer, business analyst, marketing communications manager, business architect, project manager, program integrator, program manager, account manager, data security consultant, solutions architect and trusted advisor, including:

- Thirty years of business and information technology experience including over five years as Chief Information Officer of a rapidly growing health plan.
- Proven track record of successfully strategizing, developing and implementing enterprise-level business and technology initiatives in the health care and financial services industries.
- An experienced and practiced consultant with the ability to work with C-level executives to develop strategy, assess capabilities, manage risk, and offer solutions that can be successfully implemented in an organization's unique environment.
- A proven communicator well versed in public speaking, meeting facilitation, webinars, journalism and video.
- A proven history of developing and implementing successful service offerings for a major IT consulting firms including Y2K, HIPAA Security and Privacy and HIPAA Electronic Transactions.

Currently Mr. Beauchamp is championing a capability-based approach to data quality. He is a leader in the adoption, education, and implementation of Data Sourcing as a corporate capability. He is in the process of writing a book on Data Sourcing.