The Information Life Cycle - Vital for Managing Data Quality

Danette McGilvray

Granite Falls Consulting Inc

danette@gfalls.com

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, Danette helps organizations enhance the value of their information assets by incorporating information quality management into the business. She also emphasizes communication and the human aspect of data quality and governance. Danette is the author of Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information™ (Morgan Kaufmann, 2008). An internationally respected expert, her Ten Steps™ approach to information quality has been embraced as a proven method for creating, improving, and managing information and data quality in the enterprise. Her trademarked approach, in which she has trained Fortune 500 clients and thousands of workshop attendees, applies to all types of data and all organizations. Her book is used as a textbook in university graduate programs. The Chinese translation was the first data quality book available in Chinese.

Danette helps clients solve specific data quality problems through data quality projects or incorporating data quality activities into other projects or methodologies. In addition to projects, Danette helps companies set up data quality and governance programs - formal on-going initiatives that address business needs by providing a foundation and services to sustain data quality. Her approach is outlined in her chapter on Data Quality Projects and Programs, in: S. Sadiq (ed.), Handbook of Data Quality Research and Practice (Springer-Verlag Berlin Heidelberg, 2013).

Danette is an invited speaker at conferences around the world and received IAIDQ's Distinguished Member Award in recognition of her outstanding contributions to the field of information and data quality.

"Doctor, doctor, my left arm hurts!" You race to the hospital only to have the doctor put your arm in a sling, give you an aspirin, and tell you to go home. But what if you were really having a heart attack? What would you expect the doctor to do?

You expect the doctor to make a proper diagnosis and institute emergency measures to save your life. Keeping the patient alive is top priority! After being stabilized, you expect to work with the doctor to identify what contributed to the heart attack. (Is there too much fat in your diet? Are you exercising?) You expect the doctor to correct any damage, if possible, and monitor your health with periodic tests and visits to her office.

Good health requires **committed involvement** from the patient, **effective processes and practices** (such as regular exercise, good eating habits, the right amount of sleep), and various **people** with **knowledge** (such as doctors, nurses, technicians, nutritionists) applying the right **technology** (such as echocardiograms and blood tests) throughout the **lifetime of the patient**. Good health also requires **motivation**. We have to be willing to put in the time and effort to take care of ourselves.

Likewise, healthy information requires **committed involvement** from various **people** with **knowledge** carrying out **effective processes and practices** (such as all those who create and maintain the data ensuring they do no harm to other systems and people using the data elsewhere; data quality analysts, data modelers, and information architects applying their specialized skills; business subject matter experts who know how the information is used; managers who reward attention to good data management practices) and applying the right **technology** (such as applications built with an eye to data quality, networks that move the data without corruption, and data quality specific tools) throughout the **lifetime of the information**. Healthy data also requires **motivation** – with the work tied to business goals, strategies, issues, and opportunities. It is never data quality just for the sake of data quality. Every organization needs information to make informed decisions and take effective action to provide products and services and achieve the mission of the company.[1]

---

[1] While I often use the words "business" or "company", everything in this article applies to any organization of any size: for-profit businesses, government, education, healthcare, nonprofits and charities.

The best decisions and actions are based on high quality information - meaning we can *find* the information we need (we can get to it and access it), it is available *when* we need it (it is timely and not late), it includes *everything* we need (nothing is missing), it is *secure* (safe from unauthorized access and manipulation), we *understand* it (we can interpret it), it is *correct* (it is an accurate reflection of what is happening or what did happen in the real world). Because of all these things we *trust* it when we get it and we can *use* it with confidence.

Historically, when data and information are referenced, the focus has been on technology. Technology is vitally important! No one wants to go back to the days of carbon paper and mimeograph machines. (Anyone out there remember those?) However, saying that you will have high quality information just because you have a tool is like saying that you will have good health just because you have an x-ray machine. I am not advocating less attention to technology, but I am promoting equal attention to data and information.

POSMAD Information Life Cycle* Phase	Definition	Example Activities for Information
Plan	Prepare for the resource	• Identify business objectives • Information architecture • Model and design • Business rules and standards
Obtain	Acquire the resource	• Create records • Load purchased data • Capture data • Acquire data
Store and Share	Hold information about the resource electronically or in hardcopy, and make it available for use	• Store data electronically or as hardcopy • Distribute data via networks • Subscribe and publish
Maintain	Ensure the resource continues to work properly	• Update data • Cleanse or scrub data • Match and merge, de-duplicate records • Enhance or augment data (e.g. Add GPS coordinates to an address) • Transform, parse, or standardize • Validate or verify data • Consolidate or integrate data
Apply	Use the resource to accomplish your goals	• Retrieve the data • Run reports • Analyze information • Make decisions • Complete manual transactions • Run automated job flows • Consume data
Dispose	Discard the resource when it is no longer of use	• Delete records • Archive information • Purge • Retire data

* Note: The Information Life Cycle may also be referred to as the Information Resource Life Cycle, Data Life Cycle, The Information Value Chain or the Information Chain.

Table 1: The Information Life Cycle Phases and Activities

The Life Cycle of Any Resource

Any resource - people, money, facilities, equipment, materials, products and information - must be managed from planning to disposal to make the best use of that resource and ensure quality meets requirements. The acronym, POSMAD, represents the six fundamental phases of a resource life cycle – <u>P</u>lan, <u>O</u>btain, <u>S</u>tore and Share, <u>M</u>aintain, <u>A</u>pply, and <u>D</u>ispose. Numerous activities take place within each of the phases.[2]

For financial resources, one plans for capital, forecasting and budgeting; one *obtains* financial resources by borrowing through a loan or selling stock; one *maintains* financial resources by paying interest and dividends; one *applies* financial resources by purchasing other resources; and one *disposes* of the financial resource when you pay off the loan or buy back the stock.

For human resources, one *plans* for staffing, skills, recruiting and the like; one *obtains* human resources by hiring; one *maintains* human resources by providing compensation (wages and benefits) and developing skills through training; one *applies* human resources by assigning roles and responsibilities and putting skills to use; and one *disposes* of human resources through retirement, "downsizing" or through employees leaving of their own accord. Information about both financial and human resources must be *stored* and *shared* in some manner so they are available for use when needed.

Likewise, many activities must be managed to have high quality, trusted information. Table 1 describes the life cycle phases and provides sample activities as they apply to information.

---

[2]   Many thanks to Larry English for teaching me about the universal resource life cycle (Plan, Acquire, Maintain, Dispose, Apply). See Larry English, Improving Data Warehouse and Business Information Quality by Larry English (John Wiley & Sons, 1999), pp. 200–209. I modified the names of the life cycle phases slightly from his original, added the "Store and Share" phase, and developed the acronym POSMAD as a reminder of the fundamental Information Life Cycle phases. He provided the original examples for activities within the phases for human, financial, and information resources which I have further modified.
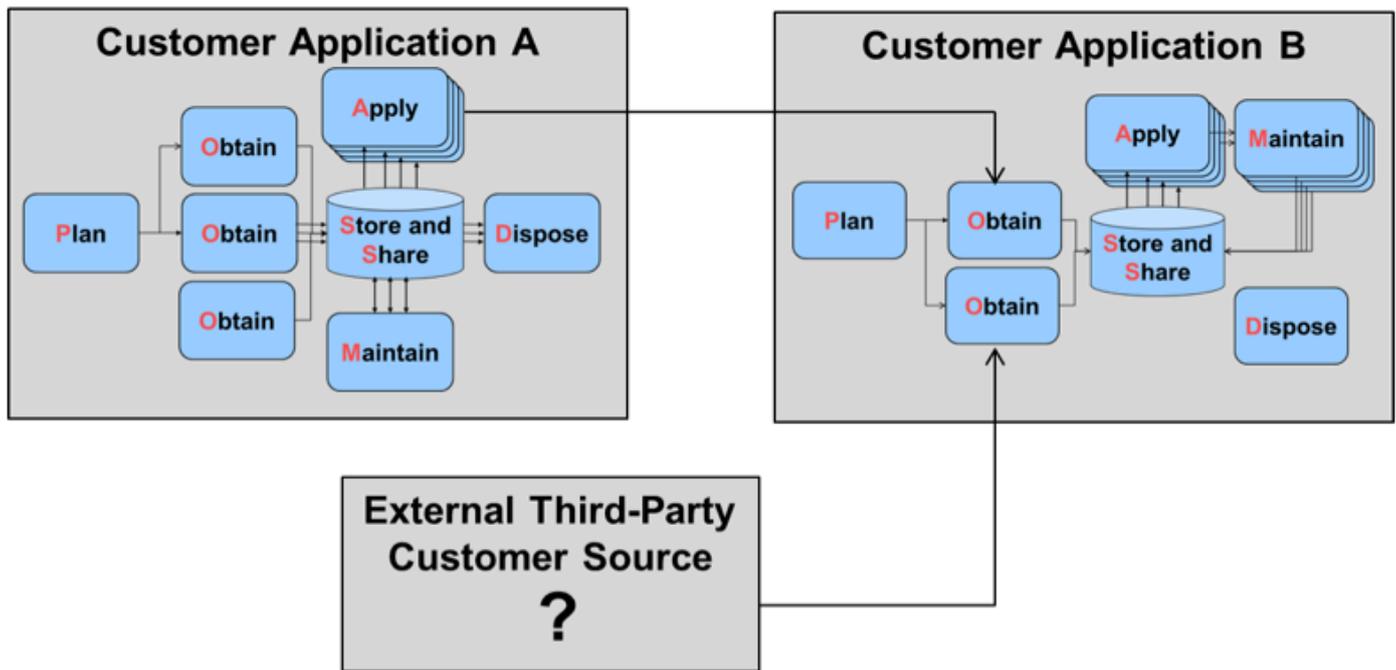
Figure 1: The Information Life Cycle is not a Linear Process

The Information Life Cycle – Not a Linear Process

I have discussed the life cycle as if in the real world these phases and activities happen in a very clear, recognizable order. This is not the case. Figure 1 illustrates the phases in the information life cycle.[3] Note that the life cycle is NOT a linear process and is very iterative.

There can be multiple ways that any piece of data or set of information is obtained, maintained, stored and shared, applied and disposed of. In Figure 1 you can see that one application's Apply is another application's Obtain; Planning is not always done in Customer Application A, Dispose is noted, but not being addressed in Customer Application B, and we don't know anything about how the information is managed for data purchased from the external source.

The information life cycles we have to deal with are much more complex than Figure 1 shows. Every organization I have ever worked with has some version of what I call a "spider web" graphic showing the systems that house the data and how they interact together. Even at a high level there are so many arrows and flows that it looks like a spider web. It is precisely because we work in such complex environments that we need to use "life cycle thinking" when dealing with our information.

Suppose you have to ensure the quality of the information used for business intelligence reports pulled from an existing data warehouse. A privacy and security initiative is also interested in understanding those with access to the information in these reports. Tracing the information from those reports through the Obtain, Store and Share, Maintain and Apply phases of the information life cycle helps provide the insight needed for both projects. (Note that you may be starting with the Apply phase and working backwards.) In some cases, the data could be purchased from an external source, then received by your company and stored – perhaps initially in a temporary staging area. The data is then filtered and checked before being loaded into the data warehouse. The data could also emanate from several transactional applications within the company—being manipulated through filters, transformations and de-duplication.

It is easy to see how the information path quickly becomes very complicated. Life cycle thinking is also critical when developing applications so you can implement a stable information life cycle and processes that will ensure quality data and information from the beginning.

To have high quality, trusted information, four key components must be managed throughout the life cycle of the information: 1) Data – known facts, items of interest to the business, 2) Processes – procedures, activities, actions, or tasks that touch the data or information, e.g. business processes, data management processes, processes external to your company, 3) People and Organizations – the business units, teams, roles, responsibilities and individuals that affect or use the data or are involved with the processes, and how they are structured, 4) Technology – forms, applications, databases, files, user interfaces, programs, code, networks, media, etc. that store, share, and manipulate the data, are involved with the processes, or are used by the people and organization.

Information is a Reusable Resource

A major difference between information as a resource and other resources is that information is reusable. It is not consumed when used. Once a product on the shelf is purchased by a customer, it is no longer available for the next customer to buy. Once materials are used to build that product, they are not available to be used in the next manufacturing cycle. What happens with information? Just because Sam runs a report on the first of the month, does the information disappear when Maria runs her report on the tenth, or when Patel accesses the information to help a customer? Of course not! When information is used, it is not depleted. The implications of this difference are important:

---

[3]   Adapted from a figure in Larry English's Improving Data Warehouse and Business Information Quality (John Wiley & Sons, 1999). Used with permission.

- *Quality is critical.* If the information is wrong, it will be used again and again – with negative results. And each time, the poor-quality information causes more cost to the company or may result in lost revenue.

- *The value of the information increases the more it is used.* Many of the costs in planning, obtaining, storing, sharing, and maintaining the information have been expended. Often with little incremental cost, the same information can be used in additional ways to help the company.

Value, Cost and Quality and the Information Life Cycle

It is important to understand value, cost and quality in relation to the information life cycle. Some key points:

- *All* phases of the information life cycle have a cost.

- Data quality is *affected* by activities in all of the phases in the life cycle.

- It is only when information is *applied* that the company receives value from it. If the information is what the knowledge worker expected, and is useful when applied, then it is helpful and has value to the company. If the quality is not what the knowledge worker needs, then that information has a negative impact on the business.

- By viewing information as a resource, you can determine its costs and its value to the business.

Why the Information Life Cycle is Important

Applying your knowledge of the information life cycle brings clarity to complex, and often messy, situations. Having a consistent view of an information life cycle, tracing how data is moved and used throughout an organization, and managing the associated people/organizations, processes, and technology will help:

- Make sense of a very complex environment

- Diagnose practices and processes and realize where breakdowns are occurring

- Develop more stable processes by taking the other key components into account

- Clarify accountabilities and identify those responsible for the various activities throughout the information life cycle

- Design better flow between and within technologies by planning for the other key components

- Enable conversations that lead to better identifying, prioritizing, and coordinating where data quality efforts are already taking place and where there needs to be additional focus

- Ensure key activities related to data, processes, people/organizations, and technology throughout the life of the data are addressed and aligned, whether operationally or in projects

- Save time in development, business impact, and root cause analysis

- Support consistently measuring or comparing the health of the data as it moves through the enterprise

While a business really only cares about the information when it wants to use it, resources should be devoted to every phase in the life cycle in order to produce the quality information needed. In practice, it is impossible to do everything at once and it is often not feasible to address all phases of the life cycle at the same time. However, a business should know enough about what happens in each phase and carefully consider how the information is being managed (and needs to be managed) in every phase in order to make informed decisions about investing in their information resource.

In Summary

Returning to our health example, no one wants a doctor who graduated from medical school knowing only how to diagnose a problem using a memorized checklist or online application for checking symptoms. Of course, it is important to know common symptoms and what they can mean, because you want the doctor to be able to quickly diagnose and administer whatever is needed to get you on the road to good health. But what if your condition isn't standard? Don't you want a doctor who is versed in the science and art of healing? Someone who can treat injury and disease? You want a doctor who knows practices that can restore health by preventing and treating illness, and has a foundation in medicine, health, treatment of diseases. You want a doctor who can look at your particular situation and apply his or her knowledge to address your problems.

Achieving high quality information also requires more than a template or a checklist. Organizations of all types need information doctors - the data quality professionals who know how to apply concepts and practices to prevent data disease and restore data quality health. Our companies need the professionals who can look at a specific situation and apply his or her information quality knowledge – fundamental knowledge – to address the problems. We need those who have a foundation in the many concepts and practices of information quality, including the information life cycle and the four key components discussed in this article. These are critical aspects of the "wellness" program for data and information with you as the information doctor. Hone your skills, apply life cycle thinking and get to work – the world needs you!

Portions of this article are from the book, *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information™*, by Danette McGilvray, published by Morgan Kaufmann Publishers, Copyright 2008 Elsevier Inc. All rights reserved.