

Guest Column | May 11, 2020

## Data Integrity In A Cloud-Based World: Regulations & Best Practices

By [Kip Wolf](#), Tunnell Consulting, [@KipWolf](#)

To begin, it is important to understand that data quality and data integrity are not the same thing.

Data quality may be defined as the general utility of a data set as a function of its ability to meet the requirements for its use. This definition includes relativity that may also be explained as bias, which simply means that context is necessary to fully interpret and understand the data. Data quality is very specific to the data set and the data itself and, if measured to be poor, may be improved through verification, transformation, and/or cleanup.



Data integrity is about trust and is as much about the supporting systems and processes as it is about the data set and the data itself. Data integrity relates to the state of the data or the sensitivity of data to external influence or change.

### Data Quality Vs. Data Integrity: An Analogy

The difference between data quality and data integrity is best described through analogy or example. Let's consider the contact list data in your mobile phone. "Home" is likely the default phone number type when entering a new contact. Most contacts in our lives likely no longer maintain a landline phone in their home, or they likely prefer their mobile phone as their primary contact. When we enter their mobile number in the default field in our contact list, do we change the record type from "home" to "mobile"? And if we do, are we consistent in the use of the phone record type? Do we change it sometimes or all the time? And are we accurate? Do we change it to "mobile" even when we are unsure whether the number is "home," "office," or "mobile"? This is an example of data quality.

Data integrity, however, can be represented by the controls imposed on the contact list by our mobile phones. If the contact list on our phones were accessible and editable by others, that could represent a risk to data integrity. But if the contact list is accessible and editable only by the owner of the mobile phone, is backed up to the cloud automatically on a regular basis, and has a record of what changes were made and when, we may consider data integrity to be rather high. Even if some contact list data records are incorrectly classified as "home" rather than "mobile," the data integrity can still be high. And if we have implemented an auto-lock feature, biometrics, or additional passcode features on our phones, we may realize even more improved data integrity.

One may argue that the example still blurs the separation of data quality and data integrity, as the incorrect classification of phone numbers is an accuracy issue, one of the five ALCOA (attributable, legible, contemporaneous, original, and accurate) principles that are foundational to data integrity. This is true, and it reflects the value in the example to show that data quality and data integrity are closely intertwined; it also provides the extension of this example to consider how this understanding shows the relationship to data quality by design.

### QbD: A Foundation For Data Quality

The concept of Quality by Design (QbD) was formally outlined by J. M. Juran and uses a systematic approach to design for quality and innovation, one of Juran's three universal processes, known as the *Juran Trilogy*:<sup>1</sup>

1. Quality planning
2. Quality control
3. Quality improvement

The three managerial processes of planning, control, and improvement make up the trilogy, with the first (planning) providing the foundation for good data quality. By applying quality planning to the data set even before the first data is created, we can ensure better data quality and improved data integrity.

The historic approach to data quality has relied on lagging indicators. We are expected to constantly consider the current regulatory and operational landscape for opportunities for continuous improvement. We follow innovations and trends in advancing technologies, but we often tend to wait to see what new technologies are implemented by other industries before we take the risk of making the changes in our own highly regulated and risk-fraught operations. We look for trends in regulatory inspections by poring over reports, observations, warning letters, and consent decree data to understand the contemporary thinking of the regulatory inspectors and related boards of health. All of this thinking may be related to the "c" in cGMP — the "current" interpretation and enforcement of good manufacturing practices and related technologies.

The problem with this approach is that it is very reactive. This is not a criticism of the continuous improvement cycle that is obviously reliant on historic data (even if it is near real time). The control and improvement processes are dependent on feedback loops for which we need to maintain such reactive processes (but should shorten the feedback and response times, a topic for another day). No, here we are focusing specifically on the planning process.

The link here to data QbD and the influence on data integrity is best summarized by the concept of improving and expanding our planning horizon. This is necessary to adjust and respond to the ever-increasing number of disruptors to our operations, such as the IoT (Internet of Things), blockchain technology, and artificial intelligence (AI). We must focus more than ever on data quality planning by understanding the intended use of the data, even extending the design considerations to the potential use of the data by interpreting the likelihood of future innovation and disruption (a more ambiguous and tenuous task).

### **Shifting To Data QbD In Each Stage Of The Product Life Cycle**

We must change the way we think. We must stop perceiving our processes as separate from the current supporting and related downstream technology. We must begin to consider and care as much about how our data will be used or interpreted as we do about how we create it. This requires that we all bear the obligation of improving our technological understanding and accept that we must continue to develop our knowledge of data-related technology. The concept of data QbD for the improvement of data integrity has application to all of us, regardless of our life sciences segment, where we find ourselves within a business or product life cycle, or in what division or function we operate. Shifting our way of working is necessary to adopt the approach for data QbD.

#### ***In Discovery Research Or Early Development***

Without constricting the creativity of research science, establish some minimum requirements for data quality and data integrity. Consider and plan for how your data may be used in the future to improve current practices and implement new practices. For example, capture data electronically and verify minimum data standards and information controls to enable more efficient transfer and adaptation of that data to downstream processes (e.g., method transfer, development activities, or regulatory submissions). We have found this to prove extremely valuable in terms of shortening tech transfer timelines, improving the effectiveness of development activities, and increasing efficiency around regulatory submission activities.

#### ***In Late Development Or Commercial Manufacturing***

Go beyond the traditional considerations for things like data structures, data storage, backups, and disaster recovery. Consider things like usability, interoperability, user experience, and the like. Consider the context in which data is created to realize how it may be misinterpreted down the line (i.e., bias). Think beyond the computer attached to the device or equipment and consider where the data will go and who will be interpreting it. Instead of just qualifying the lab equipment and relying on the IT or CSV group to validate the attached computer system, consider your obligation to define what data is being created, where it will be stored, and how it will be used for the entire life cycle of the data. Instead of simply relying on the vendor to install and perform system testing or IQ/OQ on the manufacturing equipment (e.g., bioreactors, mixing equipment, packaging machines), ask questions about data creation and how/where data will be used downstream. Draw simple data flow diagrams going beyond the individual functional area that is responsible for the equipment to understand inputs and outputs of the process to inform data quality planning. Do this in simple, common language as a business exercise and not in exhaustive detail (e.g., not a Kaizen event or detailed process mapping activity). The value comes when all parties involved understand the simple data flow to interpret and uncover valuable data quality opportunities (that can then be translated into the detailed processes for business process mapping, equipment qualification, or computer system validation of which we are so familiar).

#### ***In Supporting Operations Or Partner/Supplier Contracts***

Adopt the mentality of a service level agreement by discussing data input and output expectations for the relationship from a user's perspective. Proactively discuss and plan for specific data needs in terms of acceptance criteria (e.g., data formats, metadata requirements, metrics and measurements). Define these requirements clearly and have all key stakeholders endorse them through something as simple as email concurrence or as formal as a requirements document, specification, or contract. Think in terms of inputs/outputs, creation/acceptance, and other language that elicits the concept of user-focused service level agreements, even in the absence of such formality. Think about how the data may be used in this and other transactions. What about other partner/supplier relationships (i.e., is the data relative, shared, or sensitive)? How might the data inform other opportunities for improvement both in internal processes and in future data relationships? What happens if/when the relationship changes and how will the data be interpreted or valued (e.g., in merger or acquisition, technology transfer, project cancellation/suspension)?

### **Data QbD Demands Up-Front Planning**

The bottom line is that we must adopt more planning in our data quality activities. This is data quality by design in its simplest explanation, and we must include data integrity planning in those activities (e.g., considering how to meet ALCOA requirements). Adopting a new or revising our current operational paradigm requires a transformational change. It is necessary to do so to keep up or be competitive in our ever-changing world. A failure to do so not only decreases the value of your data assets, it presents a risk to your operational sustainability (at a minimum). It demands cross-functional thinking and enterprise risk management (in which quality risk

management should be included). This requires strategic vision and a less tactical approach to quality management, data quality measurement, and data integrity operations. Those who are up to the challenge will realize great success for themselves and value for their organizations. Those who are not will be left behind and put their organizations at greater risk.

### References:

1. Juran, J.M. (1992). *Juran on Quality by Design: The New Steps for Planning Quality into Goods and Services*. Free Press.

### About The Author:

Kip Wolf is a principal at Tunnell Consulting, where he leads the data integrity practice. Wolf has more than 25 years of experience as a management consultant, during which he has also temporarily held various leadership positions at some of the world's top life sciences companies. Wolf temporarily worked inside Wyeth pre-Pfizer merger and inside Merck post-Schering merger. In both cases he led business process management (BPM) groups — in Wyeth's manufacturing division and in Merck's R&D division. At Tunnell, he uses his product development program management experience to improve the probability of successful regulatory filing and product launch. He also consults, teaches, speaks, and publishes on topics of data integrity and quality systems. Wolf can be reached at [Kip.Wolf@tunnellconsulting.com](mailto:Kip.Wolf@tunnellconsulting.com).



## Comments

Login

There are no comments posted yet. [Be the first one!](#)

### Post a new comment

Enter text right here!

Comment as a Guest, or login:

Name

*Displayed next to your comments.*

Email

*Not displayed publicly.*

Website (optional)

*If you have a website, link to it here.*

Subscribe to

**Submit Comment**