The Importance and Role of Data Integrity in the Life Science Industry

Nine Key Steps to Getting Data Architecture Right the First Time

CONFIDENTIALITY

DATA

AVAILABILITY

INTEGRITY

VERISTA

Contents

Why Should Organizations Always Think About **Data Governance?**

Data Integrity and Its Role in a Comprehensive Data Architecture

Beyond the new government regulations and laws such as Sarbanes-Oxley, General Data Protection Regulation (GDPR), Sunshine Act and recently the California Consumer Protection Act, implementing a data governance program is a good business practice for the following reasons:

1   Ungoverned data presents an immense risk to any organization. Decisions are made on data and any inaccuracies or use of irrelevant data can be costly. Almost every executive would agree that data is one of a company's biggest assets and it should be treated as such.

2   There is a large amount of external data available to most companies which can become a huge competitive advantage for any company using it properly. The ability for a company to bring in new data sources and integrate them into their current systems requires a mature and robust data architecture and a data governance framework for managing that data.

3   As organizations mature and enterprise applications are implemented, unfortunately inefficient excel spreadsheets continue to contain critical data and with many errors. Data governance programs must address these issues.

4   In most organizations, data flows throughout the company without control or documentation. Oftentimes, system ownership for key data entities has not been defined, data is updated in multiple places and reports throughout the organization are not consistent. To overcome these challenges, some organizations have implemented a Data Warehouse solution as part of the data governance architecture.

5   Data has value, but only when it is accurate, timely, available, shared efficiently, and well defined. Companies cannot realize the full value that data has to offer until they fully govern it.

6   The law requires it. Regulations protect consumers and therefore require companies to actively protect consumer data. New laws allow consumers to ask how their personal data is being used and sold and can require the company to delete all their data from the organization's systems. Most organizations are simply not ready for this without a robust data governance program.

7   Data governance is needed to meet compliance requirements. Core to FDA and other health authorities' requirements is maintenance of all documentation around clinical trials with authorization and signoff at key steps. Authorization ensures accountability.

Data Integrity
Defined Through CFR 21 Part 11 (aka ALCOA+)

Data Integrity will be defined here as the components of the 21 CFR Part 11 and ICH E6 R2 (and soon to be adopted R3) regulations. The acronym ALCOA was part of the original guidance to ensure data integrity. The additonal 'plus' items were added by the industry to further define tenents of best practices data integrity – the cornerstone of a data governance program.

The elements are summarized below:

1 **ATTRIBUTABLE**

There must be accountability of the data. Records should be electronically signed to show ownership of the approval of the data. At the enterprise level, organizations must assign ownership for all data. Some data entities will be owned by multiple groups. For example, as a product progresses through phase 1, 2, 3 and commercial, there may be different owners for the product data and even key attributes (fields such as therapeutic area for the product). Making sure that owners are accountable for their data, even when it leaves their hands, is critical in removing risks from data.

2 **LEGIBLE**

The information must be easily understood. In addition, the data must be in a format that is easily digestible. Having standardized and accessible reference data entities will save time, effort, and many mistakes. Locate and define key entities such as patient, product, sales territory and make them accessible and secure. Don't allow updates to the reference data without updating the "golden copy" and ensure that your processes define how to tie data back to this source of truth.

3 **CONTEMPORANEOUS**

Record approval and documentation should be performed at the time of the test. The creation of a data quality framework is key to success. As organizations transform and summarize data, they should test the validity of the modifications and aggregations and store the results where they can be audited on an ongoing basis. Avoid rogue updating of data without verification.

4 **ORIGINAL**

Organizations must utilize the original paper or electronic versions of testing. Copies are not reliable as they can be updated or inconsistent with the original tests unless rigor is imposed on the copy and distribution process. It is impossible to have only one version of data across the enterprise since there are multiple copies of data for different departments and purposes. The key is to guarantee that what you started with is what you end up with. Having a disciplined data architecture with data governance will get there.

5 ACCURATE

Processes for analyzing or calculating the data should be robust and have ways to cross-check the results. Similar to many of the preceding items, data architecture and governance solutions must support the ability to take data and morph it into new information solutions across the enterprise. Cross-checking is more difficult than simply getting row counts. Understanding the calculations that are applied to data requires organizations to document them in business terms and that data owners verify the results every step of the way. The more this is automated, the more likely it will be done consistently.

The Following Elements Improve upon ALCOA for Data Integrity

6 COMPLETE

The company should keep an audit trail of any updates to the data to ensure that information has not been deleted or lost. Without a central data hub, this can be very time-consuming across an enterprise. If you are not able to implement a data hub, you will need to ensure that everyone that updates data keeps an audit trail. This should apply to non-GXP systems as well which means that Excel cannot be used to transform or change data. Furthermore, a system must audit and keep track anytime a report that categorizes and summarizes data is created.

7 CONSISTENT

Data updates should be timestamped so that the organization can recreate the sequence of events for the data captured. Taking the last item one step further, knowing the full data lineage of all data requires careful coordination. Central data hub tools make this much easier to apply across an enterprise. While this type of solution can appear to be a lot of extra work, it's work that should be done.

8 ENDURING

The information that is gathered and stored should be saved for as long as required, sometimes for decades. To endure data requires that you know where all copies are stored. Assigning data ownership and applying at least light data governance will help with this requirement. This should be applied to all non-GXP data as well (at the least, critical non-GXP data).

9 AVAILABLE

Organizations must be able to easily access the data, signatures and the chronology of events. Data is not an asset if you don't know where to find the highest quality source of information. For example, commercial data is required to gain competitive advantage from your data.

Key Steps to Getting Started with Data Governance

The Following Concepts Summarize the Key Steps in Implementing a Data Governance Program

STEP 1

Incorporate Regulatory and Legal Requirements into the Plan

All data governance programs must comply with regulatory and legal requirements. This section discusses the best practices that address general requirements you will or may be required to follow. While specific laws may involve specialized procedures and policies, the following set of guidance applies to key regulations and laws.

DATA INTEGRITY

Data integrity best practices always allow organizations to fully understand their data. Any updates, additions or deletions to data must be tracked or audited. The company must ensure that when data is moved from one system to another, or transformed, those transformations are the only thing that happens to the data. Procedures must be put in place so data remains accurate and complete as it flows through the organization.

DATA CLASSIFICATION

Regulatory and legal bodies understand that requiring companies to proactively manage and protect data is expensive and time consuming. Because of this, many legal requirements focus on sensitive data that presents the greatest risk to the people the company holds data for. Information that identifies individuals such as name, birth date, gender and address is particularly sensitive information and must be classified as one of the highest risks to an organization.

DATA PRIVACY

Data privacy requirements focus on ensuring that only authorized people view data elements for lawful and appropriate use. Best practices apply privacy rules to data that are classified at the attribute level such as birth date, address, email, adverse event, pre-existing conditions, etc. When information about a subject involves data attributes that are classified differently, the organization must ensure that protection is enforced at the attribute level. From this, storage, transformation, and security policies at the classification level are applied.

DATA SECURITY

The organization is responsible to protect data from theft, invalid viewing, and accuracy. Data should be classified to ensure that it is known which data is most at risk and presents the greatest harm. Furthermore, stringent data security is an imperative across all data elements as they flow through the organization. The company will have different requirements for data reporting and sharing, so it is critical that all requests for data sharing focus on data classification and the security for that group of data.

DATA RETENTION

While all data is valuable, the organization must have policies and procedures that deal with data retention for information that is no longer required legally. Understanding the data subject's desires to opt in or out is part of the decision-making process for keeping data beyond it is necessary use.

STEP 2

Take and Re-take Inventory of Your Data

Your next step is to determine your organization's data assets. Each department needs to document EVERY data asset, including spreadsheets, presentations, word documents, lab notebooks and notebook type applications, emails (as there is a great deal of valuable data stored in many organization's email systems), small and large applications, individual file folders on employee desktops, and shared folders (including Dropbox, SharePoint, OneDrive, Google Drive, and network shares). This inventory should be updated on a regular basis.

It will become apparent how ungoverned and widespread the company's data assets are. Not everything can be identified at first, so aim to find key data assets first. The results should be audited on an ongoing basis to ensure that the inventory is accurate and up to date. The process of inventorying all aspects of data assets is a significant undertaking that should not be taken lightly.

STEP 3

Establish Data Ownership

Your next step is to determine your organization's data assets. Each department needs to document EVERY data asset, including spreadsheets, presentations, word documents, lab notebooks and notebook type applications, emails (as there is a great deal of valuable data stored in many organization's email systems), small and large applications, individual file folders on employee desktops, and shared folders (including Dropbox, SharePoint, OneDrive, Google Drive, and network shares). This inventory should be updated on a regular basis.

It will become apparent how ungoverned and widespread the company's data assets are. Not everything can be identified at first, so aim to find key data assets first. The results should be audited on an ongoing basis to ensure that the inventory is accurate and up to date. The process of inventorying all aspects of data assets is a significant undertaking that should not be taken lightly.

STEP 4

Trace Data Movement Throughout Your Organization

Once the organization has located data throughout the company and assigned ownership, it can now be traced as it flows throughout the company and to external entities. Each of these flows can now be managed, removed, or altered based on the perceived value of the data movement. Since the company has assigned ownership to data, it can begin to govern transport. In some cases, the organization will add data movement to new departments or applications to increase the value of the data assets. Each data movement should be assigned either value or risk to assess if it should be retained.

STEP 5

Track Data Usage and Downstream Business Rules

Once data has been inventoried and its programmatic movement throughout the company has been documented, we can now track end user data usage/sharing. This allows organizations to do change impact analysis if changes are made upstream. It also allows them to determine reporting and analytic business rules for the data owned by a user/department. They will want to document these business rules as they will have a potential material effect on decision making. Furthermore, an understanding of business rules will aid discussions when disagreements on reporting arise (and they always will).

STEP 6

Share Data Securely and Effectively

One of the best practices of data sharing comes from government entities. Since government agencies are separate in their organizations, they typically house their own data solutions. When asked to share data, they will require Data Sharing Agreements. These agreements dictate how the data will be used and provide guidance (and rules of engagement) on further sharing of the data. Most data sharing agreements do not allow the receiving organization, for example, to share the data with any other group. This may be overly burdensome for an organization, but simply thinking about agreements between departments and how they can use data is a valuable exercise. Some companies will have Data Usage and Sharing guidelines that all key users must sign.

Once the system owners have been identified over the life cycle of the product (R&D vs. Clinical Trials vs. Commercialization) and the set mechanisms or processes to track data usage, the company can begin to craft the solution for sharing data. The first step is to identify a Chief Data Security Officer. This could be the CIO, one of their reports, or could be someone in the business. While this role typically falls under corporate IT and acts as the data custodian for the entire organization, it can exist wherever it makes sense. It is important, though, that the position has the appropriate amount of authority in the organization to authorize changes on how data is stored, moved, and shared.

The most critical aspects to sharing data safely and effectively are:

✓ Create data sharing agreements that provide guidance on using data

✓ Classify the data being requested for sensitivity and certification of validity

✓ Assign a data owner that can approve the request

✓ Provide the controls necessary for sharing that classification of data

✓ Keep an audit of the request

✓ Don't be afraid to share data, it is an enterprise asset; however, it must be protected

STEP 7

Ensure that Data Is Accurate

One of the biggest investments the organization can make is the creation of a centralized data hub. The company may choose to implement an active data hub where systems are connected and updates in one system propagate to others as needed. Reporting is typically performed as a federated solution where all the systems contain the information, and the central hub collects and disseminates it. These systems are very complex and can be expensive to optimize. They are typically supplemented with a passive solution such as an Enterprise Data Warehouse.

A data quality framework should be part of building this architecture and will report on key entity and attribute values as they migrate throughout the company. The availability of data quality metrics is critical to business decision-making with advanced analytics. For example, if it is known that the gender field only has values 30% of the time, the decision to use it in analysis will be very different than if it filled in with values 100% of the time and has been adjusted to current day thinking about gender beyond male and female (assuming the detailed breakdown is of value to the decision making).

STEP 8

Actively Look for Competitive Advantage from Data

Once a set series of processes, procedures, and a data infrastructure are in place, the company will start to gain efficiencies, make more informed decisions and be in a better state to comply with current and future laws and regulations. The organization can start to identify where the data is unique in the market and where it may have value beyond its current use. When data architecture and governance programs are matured, an organization can gain competitive advantage. Most organizations are in the awareness phase and playing 'catch-up' as they are in aggressive growth phases.

Because of the potential advantage of a data governance program, there are typically ways of improving many core processes. If R&D is tied to Commercial, the company can create a feedback loop to better inform the potential pipeline. If the organization needs to purchase or acquire data on current drugs in their pipeline and combine it with sales and marketing information from Commercial data, more informed decisions can be made on where to focus investments. Advancements in Artificial Intelligence look at current products and potential re-use. The COVID-19 pandemic has accelerated this to shorten the time for vaccines, mitigation of the virus' effects and testing.

There are many ways to take advantage of the investments in a data governance program. Assign working groups that have colleagues across the organization and focus on how to best leverage the data the company has or should acquire.

STEP 9

Continue Improving Data Governance Your Way

Most data governance programs fail because they are overly bureaucratic, and the benefits are not worth the effort (at least in the eyes of the key stakeholders). Take the above advice and make it your own. What is key is to understand where your data is, who owns it and who is using it. Seek help in defining your data governance program, but make sure it's yours and will work in your organization. Data is your most valuable asset and is well worth the time and effort!

Verista is a leading business and technology consultancy firm that provides systems, compliance, validation and quality solutions to life science companies enabling them to improve health and improve lives.

We help clients solve their most critical and complex challenges across the GxP lifecycle, from preclinical and clinical to commercialization, manufacturing and distribution - bringing together decades of knowledge, the most advanced engagement platforms and transformative technologies. This allows clients to benefit from the ease, efficiency, and trust that results from working with one partner who excels across specialties.

Verista launched in 2020 by bringing together three trusted firms: Clarke Solutions, Covex LLC, and Acquire Automation. Verista acquired Hawkins Point in 2021.Verista's clients trust the company's 500+ experts to deliver consistent, safe, and high-quality results across the product development lifecycle.

For more information, visit Verista.com