Data Governance

AN EXECUTIVE'S SURVIVAL GUIDE

smarty

Sneak Peek

Introduction

Ready to take control of your organization's data?
This eBook will help you understand key elements of data governance, including the employees, buy-in, and software you'll need. We'll get you ready to start hiring and structuring—or restructuring—your organization in a way that allows you to manage, update, and merge data more efficiently.

What is data governance?
Data governance is a set of standards, roles, and accountabilities that help organizations break down information silos and unify data. The data governance system provides businesses with rules and processes to ensure data quality, security, and efficiency, improving data intelligence.

What are the principles of data governance?
Data governance has eight important principles:

- Integrity
- Transparency
- Audibility
- Accountability
- Stewardship
- Checks & Balances
- Standardization
- Change Management

All successful data governance programs include these principles and help stakeholders come together to resolve any data-related conflicts at the organization.

Data Governance Framework & Examples

What is a data governance framework?

A data governance framework creates a single set of rules for collecting, storing, and using data in your business. Also referred to as data modeling, a data governance framework makes it easier to streamline and scale core governance processes. No matter how quickly your data volumes grow, this framework allows you to maintain compliance, guide decision-making, and support collaboration.

Why does a data governance framework matter?

A data governance framework allows your business to define and document standards, accountability, and ownership. This framework is crucial to your data governance program because it outlines the most critical components necessary for overall success.

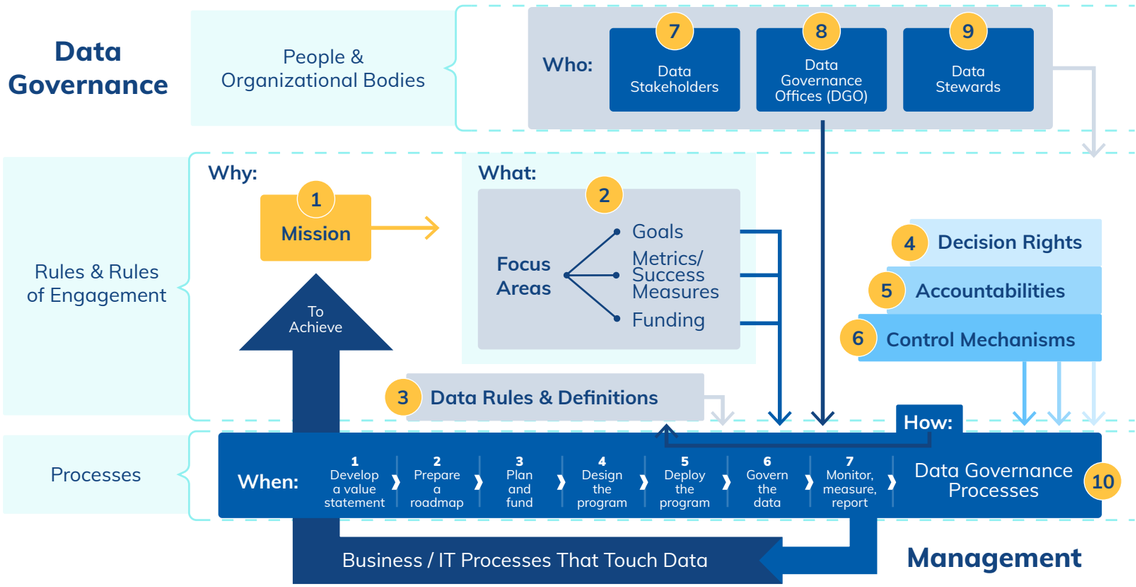
Consider this: 67% of respondents to a Profisee-sponsored Harvard Business Review Analytics Services report say that data governance is vital to achieving high-quality enterprise data.[1]

If your goal is high-quality data, you'll want to start with a solid data governance framework. So, where do you begin? Keep reading! We'll explain that in the next section, along with a visual example if you're someone who likes to see exactly how everything comes together.

What is an example of a data governance framework?

You may wonder: *"How do I create a data governance framework?"*

This template from the Data Governance Institute will be helpful.[2] While the term data modeling is often used interchangeably for data governance framework, it also can be used to describe the visual representation of that framework. Let's take a look.



Definition:

Data Governance is the exercise of decision making and authority for data-related matters.

It's a system of decision rights and accountabilities for information-related processes, executed accordina to agreed upon models which describe who can take what actions with what information and under what circumstaces, using what methods.

Processes for governing how data is used, and when, and by whom

1. Aligning Policies, Requirements & Controls
2. Establishing Decision Rights
3. Establishing Accountability
4. Performing Stewardship
5. Managing Change
6. Defining Data
7. Issue Resolution
8. Specifying Data Quality Requirements
9. Building Governance into Technology
10. Stakeholder Care and Support
11. Stakeholder Communications
12. Measuring and Reporting Value

The DGI Data Governance Framework

Here are the 10 data governance program components—shown in the template on previous page—that you'll want to identify for your business:

Mission & Vision

Here you'll identify the mission and vision of your data governance program. What are you looking to achieve? While your mission might sound dry, it should resonate with participants *and* stakeholders. Rather than being dry, describe your vision using rich and compelling language—this will inspire your stakeholders and help them envision possibilities and set data-related goals.

Goals, Metrics, & Funding Strategies

Be sure to include SMART (Specific, Measurable, Actionable, Relevant, and Timely) goals and metrics. These can act as measures of your data governance program's success. In addition, explore your funding options with your key stakeholders.

Rules

These are the data-related policies, standards, compliance requirements, business rules, and data definitions that you'll want to outline. Are there any existing rules? What new rules need to be created? What gaps and overlaps need to be addressed with these definitions?

Decision Rights

Who gets to make decisions about rules? The rights of who gets to make decisions must be identified for any data-related decision, including when new rules are created. Before any decisions are made, make sure you establish the decision rights for your data governance program.

Accountabilities

Who is responsible for taking action and when? Sometimes activities don't correspond to clear departmental responsibilities, so your data governance program should define accountabilities according to the different roles.

Controls

Data breaches are far too common today, so data is constantly at risk. Data-related controls are what operate your risk management strategies. Controls may be preventive or corrective. Either way, it's important to clearly identify the controls in your data governance program.

Data Stakeholders

These are the people who affect and will be affected by data-related decisions, including certain business groups, IT teams, Data Architects, and DBAs. Your data governance program must address their expectations, whether that's a consultation before decisions are formalized, or just a status update after decisions are made.

Data Governance Office

This group runs the data governance program. The Data Governance Office (DGO) includes the Data Governance Committee, Data Governance Manager, and Chief Data Officer. Depending on the organization, some of these roles may overlap. Some organizations also choose to forgo a formal DGO structure. Instead, they choose to employ individuals who perform the functions that would typically be assigned to a DGO.

Data Stewards

The Data Stewardship Council is made up of data stakeholders who make data-related decisions together. Large organizations often choose to structure a hierarchy of stewards, breaking them into teams or working groups. The data stewards set policy, specify standards, and craft recommendations.

Data Governance Processes

The processes—or methods—used to govern data should be standardized, documented, and repeatable. These processes include aligning policies, requirements, and controls, establishing decision rights, establishing accountability, performing stewardship, managing change, defining data, resolving issues, specifying data quality requirements, building governance into technology, stakeholder care, communications, and measuring and reporting value.

As you can see, a lot of detail goes into data governance policy. Now that you know what a data governance framework could look like, we're ready to take a deeper look at some topics and data governance best practices that will help to make your organization's program a success.

Data Quality & Data Management

What is data quality?

Data quality measures how fit data is to serve its intended purpose. Data quality is one of the most important aspects of data governance. First, let's look at what makes good quality data.

Here are the 5 characteristics of good data quality:

1. **Accuracy**
   Does the data accurately reflect an event or object that's described?

2. **Completeness**
   Does the data fulfill certain expectations of comprehensiveness at the organization? Is there enough data to draw meaningful conclusions?

3. **Consistency**
   Are there two data values—retrieved from multiple and separate data sets—that conflict with each other? Note: even if the data is consistent, this alone doesn't imply that the data is correct.

4. **Integrity**
   When structurally testing data, is there compliance with the organization's data procedures? This compliance is also known as data validation. Data integrity shows that it has no unintended errors and corresponds to appropriate data types.

5. **Timeliness**
   Is your data ready when users need it?

Next, let's look at metrics to measure whether the data is high quality.

Here are 5 data metrics often used to help organizations measure their data quality efforts:

1. **Ratio of Data to Errors**
   You can find this by tracking the number of known errors within a data set corresponding to the actual size of the data set.

2. **Number of Empty Values**
   Counting the number of times an empty field exists within a data set will reveal the number of empty values, which indicates missing information or information recorded in the wrong field.

3. **Data Time-to-Value**
   How long does it take to gain meaningful insights from a data set?

4. **Data Transformation Error Rate**
   How often does a data transformation operation fail?

5. **Data Storage Costs**
   Storing data without using it can often indicate that the data is of low quality. However, if the data storage costs decline while the data operations stay the same or continue to grow, the quality of the data is most likely improving.

Since data quality impacts the long-term success of an organization, it should be a top priority. Now, let's look at bad data, which can be costly.

Where does bad data come from?

Unfortunately, bad data can come from anywhere. Here are a few examples:

- The customer or user may input their information incorrectly.
- A computer glitch may alter the data.
- Autocorrect can cause errors.
- Falsified information is always a possibility.
- Sometimes, information that was at one time factual becomes out of date years later.
- The content may be accurate, but it's considered bad data; if the entry doesn't adhere to proper formats and standards.

We want to avoid collecting these types of bad data. That's where data management comes in.

What is data management?

Data management is another one of the most critical components of data governance. Data management facilitates data control and flow—from the creation of the data to processing, accessing, and deleting it. Put simply, data management is the upkeep of records, information, and data.

Sometimes data management is referred to as data administration, data resource management (DRM), or even master data management (MDM). While there are several different names you'll come across during your data governance journey, these terms all refer to the process of managing your business's data.

What are the benefits of data management?

In addition to the problem that low-quality data creates for businesses by making it difficult to decipher, bad data is unusable.

- Here's what data management can do to benefit your organization:
- Improve the function of your business
- Use resources efficiently
- Provide a competitive advantage
- Supply high-quality business leads

Let's use address data management as an example. Address data management includes checking the addresses for accuracy, for missing information, to see if they are up to date, and for duplicate entries.

A good data management tool will help with data cleansing (which we'll talk more about later) and data profiling. For now, just know that data cleansing corrects data mistakes like unknown data types, duplicate records, and substandard data representations. Data profiling is the process of data cleansing. Later, we'll also look at tools to help ensure your organization's data is good and that bad data is fixed, updated, or removed.

Data Governance Roles & Responsibilities

For successful data governance, many different roles must work together. In addition, each of those roles has many specific responsibilities. We've already mentioned a few of these teams and roles, but now we'll take a closer look to see how your organization's different data governance roles might look. Remember that you can customize roles to fit your organization's unique needs and size, whether overlapping roles or hiring additional employees where required.

Here are several examples of commonly used data governance roles and teams, along with some of their primary responsibilities:

Chief Data Officer

The Chief Data Officer (CDO) is a senior executive responsible for overseeing the entire data governance program. Their duties typically include securing approval and funding for the program, hiring key staff members, and leading the program.

Data Governance Manager

This program manager leads the data governance team. This role may also be the CDO. Or if there is no CDO, organizations may decide to hire a manager to serve as the lead.

Data Governance Committee

This is a team of business executives and relevant parties. The committee is responsible for making policies and standards—which are then enforced by the data governance manager and team—along with resolving disputes and amending policies as needed.

Data Stewards

These are team members specializing in specific domains and data assets—the Data Stewardship Council. A data steward can have experience with IT or business. Their responsibilities include overseeing data sets and ensuring that policies are properly implemented and compliance is achieved.

Data Quality Analysts and Engineers, Data Architects, and Data Modelers

These roles work with the governance committee and data stewards to fix data errors and track data quality metrics.

Business Users and Analytics Teams

These people must be trained on the organization's data governance policies and standards to ensure they use data correctly—and avoid using data in improper or erroneous ways.

Remember to tailor these roles and governance teams to what works best for your business. What if your team is smaller? That's okay! Some employees may wear multiple governance hats, which may work well at a smaller organization with different business goals and needs than a larger organization.

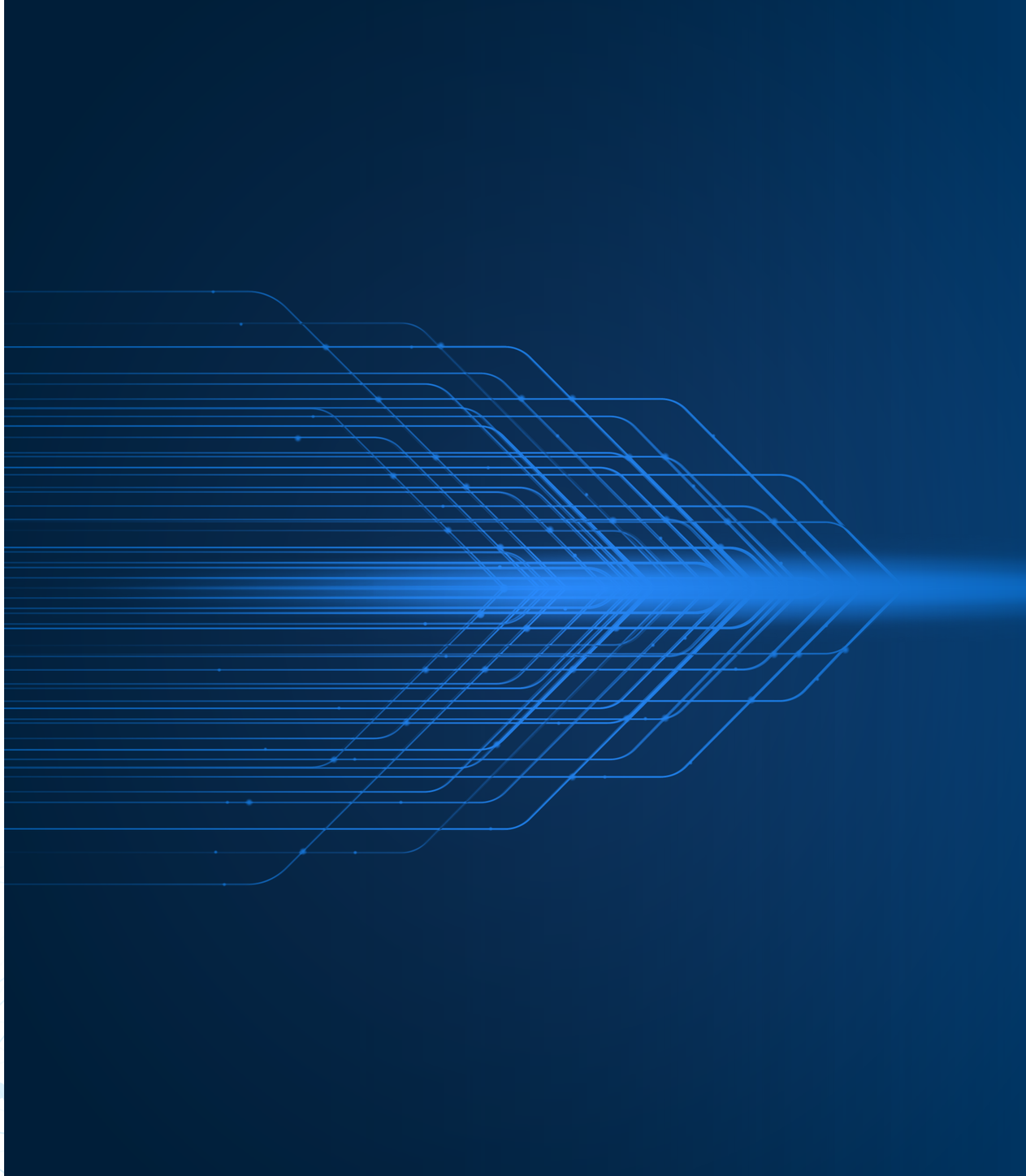
What Is Data Cleansing & Enrichment?

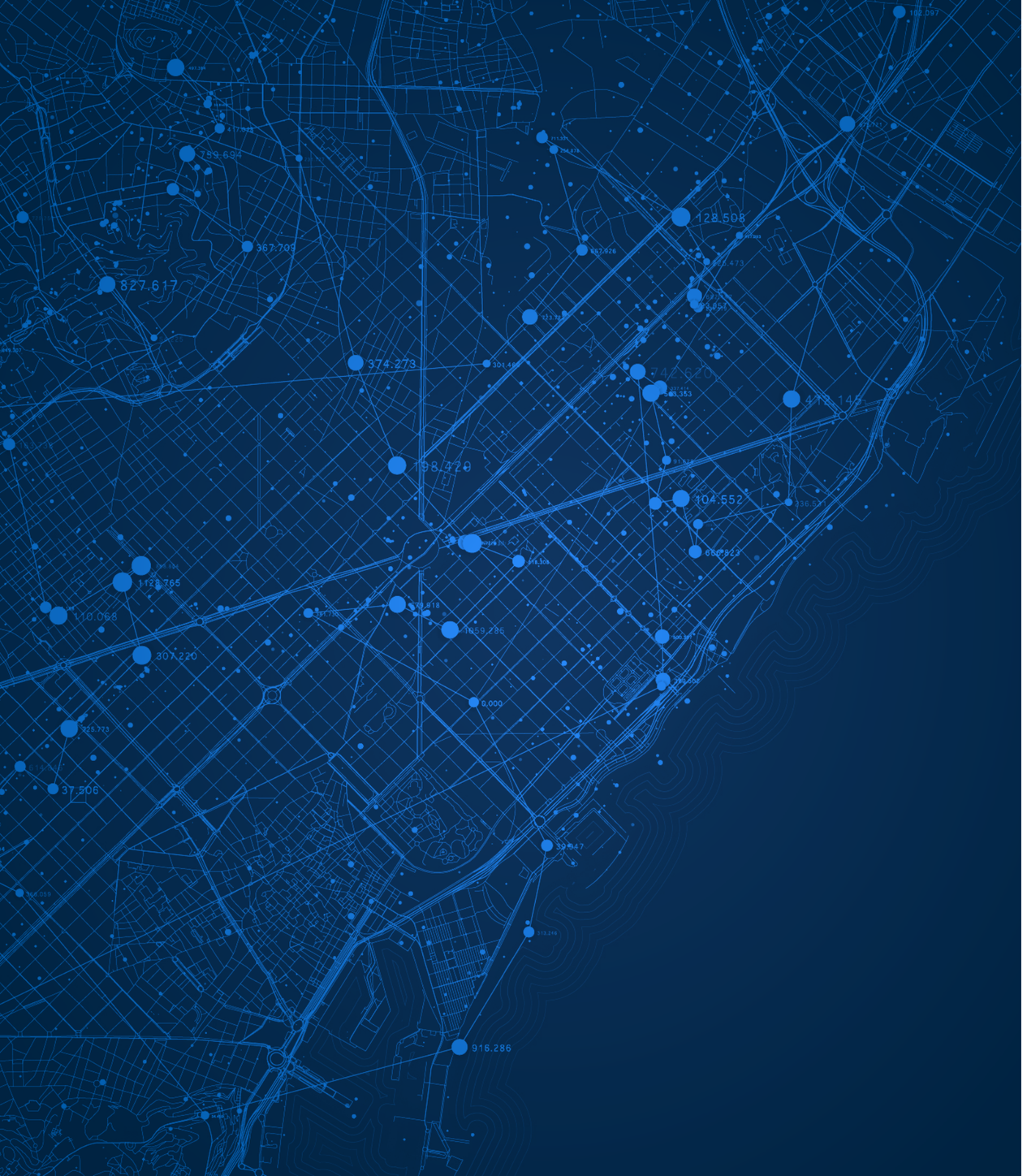
Data cleansing finds and removes corrupt or inaccurate records from a set of data. Cleansing allows you to identify which contacts in your database are bad data—in other words, data that is incomplete, incorrect, inaccurate, or irrelevant.

Data cleansing is the first step in cleaning up dirty data. But what happens to all that bad data? That's where data enrichment comes in.

Data enrichment is the process of enhancing, refining, and improving raw data. Another term you may see used for this process is data appending. When referring to data cleansing of contact data, the terms contact appending or contact enriching are also common.

During your data governance journey, you'll want to cleanse and enrich (or append) your business's data to have the most complete, up-to-date information. Data cleansing and enrichment are essential to keeping high-quality data. There are many tools available to help with this.





How can Smarty help?

When it comes to address data, Smarty's address validation tools are great for cleaning, standardizing, deduplicating, and enriching address data in existing databases, along with cleansing, standardizing, and enriching data as new information is added by customers and employees using address autocomplete.

So, Smarty helps with address cleansing and enrichment for both existing AND new data. Here's how it works:

Address Standardization, Deduplication, and Validation

Similar address data might enter your customer database for a few reasons. For example, a customer might use several marketplaces to purchase from your company. Or a customer may cancel their account and later use a different address to create a new account.

Customer overlap can even happen if your company acquires another company. Smarty's address validation software helps you delete, merge, or flag duplicate addresses—all while standardizing those addresses to match the address format of the local postal authority and validating that the addresses are accurate.

Address Enrichment

Address data enrichment takes the known data and merges it with third-party data services. Then, third-party data is appended to the known data to improve business intelligence.

Smarty's address validator takes your address list, returns the addresses in a standardized format, verifies that the addresses are real and mailable, and enriches the addresses with 40+ associated data points.

Good news: Smarty's address validation APIs are lightning fast. For example, the bulk address validation tool can validate, standardize, and append data for 75,000+ addresses per second.

Why Is Master Data So Important?

What is master data?
Master data is all the data that's critical to the operation of your business. This data is usually shared across the organization. In addition, multiple departments depend on this master data in order to make decisions.

Here are the 4 common characteristics of master data, no matter the type:

1. **Master data changes occasionally.**
   While master data tends to change less frequently than other data, sometimes it does change. Data sets that never change are not usually classified as master data.

2. **Master data is more complex.**
   It usually includes large data sets that are more complex with multiple variables—instead of similar data that could simply be tallied or counted. Plus, to keep the information up to date and accurate, master data requires processes.

3. **Master data is valuable.**
   Master data is important because it's used by a company repeatedly. This data is essential to an organization's daily operations and analytical decision-making.

4. **Master data is non-transactional.**
   While master data doesn't usually include transactional data, it can be part of the transaction process. Examples include data describing the customer, product, or point of purchase.

   Let's look at an example of master data.

   One of the most common examples of master data is customer information—names, phone numbers, and addresses. Customer information is less volatile, but it needs to be updated occasionally if a customer moves or changes their name. While it's not transactional data, customer data provides a way for the organization to connect with its customers. So if the customer information is wrong or outdated, this can cause a missed opportunity.

How do you avoid these types of mistakes? Consider establishing a master data management solution.

Master data management refers to the tools and procedures that a business uses to establish uniform data. Objectives include organizing, centralizing, and updating the master data specifically. Master data management does not have the rest of the organization's data.

Your organization may need master data management if it's very large or complex, distributes data often, or goes through frequent mergers or acquisitions. These situations present data challenges that a master data management solution can solve.

Establishing Data Security Policy & Data Security Standards

What are data security risks?

Data is always at risk for attacks and misuse. And lack of security is an increasing threat. According to Forbes, in 2019 phishing attacks rose to their highest level since 2016. In addition, the time it took to detect and resolve data breaches also rose that year.[3]

When many companies shifted to remote work in March 2020, phishing attacks increased by 667% since the end of February, as noted by Security Magazine.[4]

The increase in data breaches in recent years is no surprise to most people. However, the costs of a data breach can be detrimental to a business.

According to IBM Security analysis of the research data compiled by the Ponemon Institute, from 2020 to 2021, data breach costs rose from $3.86 million to $4.24 million.[5]

Data breaches are a huge issue. Most data leaks start as a result of improper data access permissions. That's why it's essential to establish a data security policy and data security standards as part of your data governance program. Proper security ensures appropriate access to data while maintaining privacy.

What are the methods for data security?

Safeguarding data will look a bit different in every organization, depending on how the policies and processes are defined. However, three commonly used primary software methods are: the classification, discovery, and de-identification of personally identifiable information and other sensitive data.

Here's a closer look at the 3 methods for data security:

1. **Data Classification**
   To find and protect specific data at risk, you should define the data in named classifications—categories or groups. For example, data can be classified by its name and attributes, like social security numbers. You can name data class groups by their location. Or, if the locations of the data class groups are unknown, they can be named by search methods—how to find them.

2. **Data Discovery**
   Data discovery refers to the process used to streamline locating an organization's data. For example, you can execute search functions for data classes to find sensitive data. Data discovery makes it easier and quicker to find secure data when needed.

3. **Data De-Identification**
   Data de-identification sounds complex, but it's just the process of removing identifying information from data. A form of dynamic data masking, data de-identification breaks the link between the data and the initially associated person.

The software you choose for your data governance program should help you with data classification, discovery, and de-identification to keep your organization's data secure.

Keeping Your Data Secure With Smarty

With Smarty, your data is securely encrypted. You'll connect to our service using HTTPS to take advantage of industry-grade encryption. Smarty's servers are fully automated, meaning no one will see the data you send to us. And since our database already has every address the US Postal Service has, your addresses aren't new information. We keep the names and other information associated with any given address secure—this is the important part. To avoid exposing anything sensitive or private to our secure servers, you can submit a unique identifier with the address, city, state, and ZIP.

Plus, if complete anonymity is needed, Smarty allows you to take steps to help anonymize your requests. We're committed to private and secure communications on the internet.

Any data governance software you choose should use these methods to ensure your data is kept safe. In the next section, we'll look at popular data governance tools that also take the measures needed for data security.

Types of Data Governance Tools

What is a data governance tool?

Data governance software ensures that regulatory standards are met, which then improves security and organization. It will outline and implement data quality control guidelines, increasing employee efficiency. It's important to note that you can also find software for master data management and data quality that has tools to build data governance processes, even if the software isn't dedicated to the purpose of data governance as a whole.

To be considered a data governance product, the software must facilitate data lifecycle management, provide tools for outlining and implementing governance strategies, allow the enforcement of standards and compliance requirements, and provide recommendations to improve governance processes.

Smarty Address Tools

If you need to clean, standardize, deduplicate, and enrich both existing address data along with new address data, Smarty's address cleansing tools are an excellent choice. The Smarty Address Validation APIs are fast, secure, and an excellent choice when you're looking to improve your business's address data quality—an important part of data governance.

And since address data is just one part of the big picture of data governance, you may want to pair an address cleansing tool like Smarty's with some type of data governance software. But where should you start in your search? Consider the following questions before and during the selection process.

Questions to Ask When Choosing a Data Governance Tool

- Have you established clear goals to narrow down the desired data governance tool?
- Is your data governance framework ready to use as a guide?
- What is your desired budget?
- Does your IT team have any restrictions or controls on new technology, such as cloud-only solutions?
- Do you need enterprise data governance, or is a tiered solution an option?
- Which vendor has the best fit for your requirements?
- Have you included sufficient business user involvement in the selection process?
- How long will the product take to implement?
- Does the initial implementation make sense for your business, or is it too complex?
- What training is given for the data governance team and business users?
- What do the product support, bug fixes, and enhancements look like?

Keep these questions in mind while you compare options. Here are several popular data governance software providers with top-rated data governance tools (listed in alphabetical order) to consider as you begin your search: Collibra, Claravine, Egnyte, Oracle, SAP, and Segment.

These data governance tools should help you start your search. However, you may find that your business requires something a bit different than what these tools offer. If that's the case, here are more data governance providers to look into as well: Alation, ASG Technologies, Ataccama, Atlan, Avo, Erwin, Immuta, IBM, Informatica, lyftrondata, OvalEdge, Securiti, Semarchy, Talend, and Transcend.

Remember, the best tool is going to be one that's the best fit for your specific business needs and the data governance strategy that you've outlined. As you select a data governance tool, be very clear and specific about your requirements and ask each provider questions about how their tool meets those requirements.

Conclusion

In an O'Reilly survey on the state of data quality in 2020, more than 60% of respondents said that the most common data quality issue was "too many data sources and inconsistent data." And just under 50% of respondents said that their most common problem was "disorganized data stores and lack of metadata"—which is fundamentally a data governance issue.[6]

The solution to these data quality issues? A proper data governance program with standards for business rules and data management will help to ensure high-quality data and effective governance.

With the knowledge in this eBook, you're ready to start a data governance policy at your organization. You can use everything you've just learned to create a data governance framework and appoint employees into specific roles where they can follow quality measures and make informed decisions.

Plus, you can apply what you've learned about data cleansing and enrichment, the value of master data, and how to establish a data security policy that will keep all that data safe. Choosing data governance software that's the right fit for your organization will be incredibly helpful.

We wish you the best success on your data governance journey. And if you're ready to clean up your address data with Smarty, you know where to find us.

Sources

1. 2021 Harvard Business Review Analytic Services Pulse Survey, Sponsored by Profisee
2. The DGI Data Governance Framework
3. When It Comes to Data Breaches, Hindsight Is 2020
4. Coronavirus-Related Spear Phishing Attacks See 667% Increase in March 2020
5. IBM Security Cost of a Data Breach Report 2021
6. The State of Data Quality in 2020

smarty