

WHITE PAPER

The 4 Pillars of Data Governance

OCTOBER 2021



Your CEO sends you the following email...

Hi - A friend showed me a very nice dashboard his business uses to track their key metrics. Given all the charts I see in the staff meetings, I'm sure we have all the data it takes to build one of these. And I can't imagine it's too hard. Can you whip up a basic dashboard with data from finance, sales, and marketing, maybe even manufacturing, so we can talk about it next week?

Thanks

You don't know whether to laugh or cry.

You're considered the data-minded executive, well known (but not always appreciated) for asking, "Before we draw any conclusions, what does the data say?" Now, in this sprawling organization, you're supposed to collect a bunch of scattered data and "whip up" a dashboard with the key metrics—which, by the way, have been a subject of much debate lately. You understand what the CEO is really after—a way of getting and staying informed using data everyone can trust so the team can make better decisions—but you can't do that with a couple of charts. This calls for a real strategy, not to mention a culture change.

You need to phone a friend—and fast.

If we happened to be that friend, we would tell you that we have heard versions of this story countless times over the years. There is a natural tendency to focus on the pretty dashboards without thinking about what goes into creating and maintaining them—the data governance that ensures you know and can trust what the data says. Fortunately, you're in a better position than most; the demand came from the top. You can use that opening to set expectations and get the CEO's support for a true data governance strategy. We would tell you this is an exciting opportunity to make a strategic contribution.

After talking to your friend, you decide to use this dashboard concept to persuade the CEO to make this a priority. That will ensure the other department heads give you the time and data and you need. You start drafting an email the CEO could send to the executive team to promote this project:

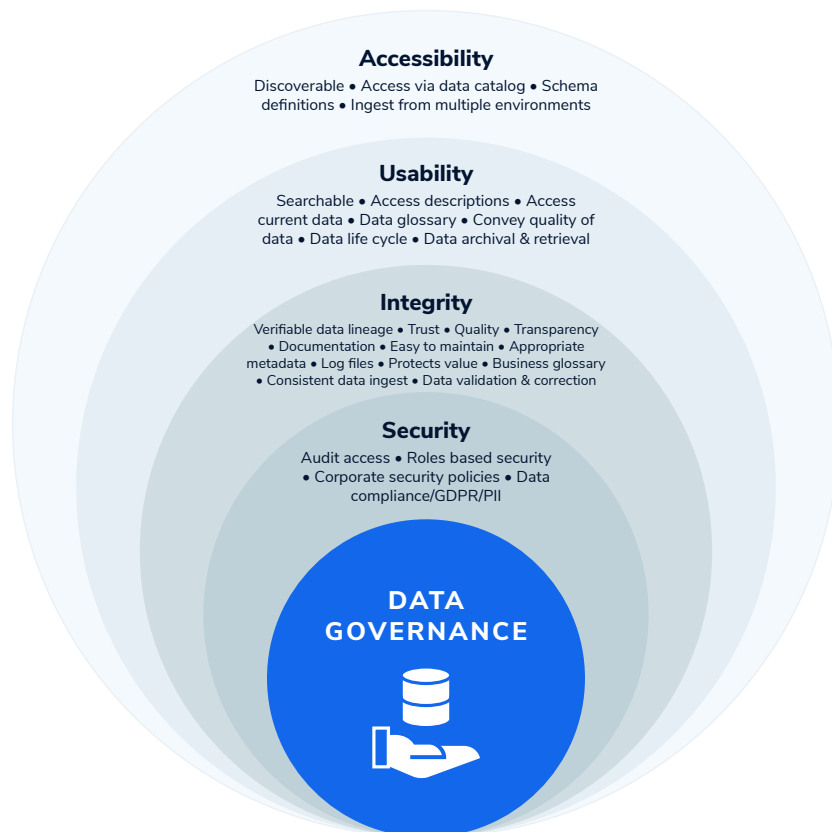
As we converge on a set of key metrics for the business, we are going to begin creating a long-term data governance policy to collect, track, and maintain those metrics in a usable form. For those who don't know, data governance is how we manage the accessibility, usability, integrity and

security of the data. To have a solid, workable data governance policy we'll need a standing oversight committee, policies and procedures regarding data standards and usage, and a project plan for implementing everything.

Rereading your first draft, you realize you have to convey the value of data governance before getting into the details. You will need their support to get all this done, after all. You try another approach:

For those who don't know, data governance is just a formal term for the infrastructure that ensures our data is properly managed, beneficial, and trustworthy. In other words, it's the policies, processes, and technology we use to leverage the full value of our data.

That's enough at the conceptual level, but to get buy-in you'll have to explain what's in a governance plan and how it helps them. That begins with the four pillars of data governance: **accessibility, usability, integrity and security**. Without these, your data will be worthless. It may be tricky to fit in an email, but you need to explain the value and importance of these four pillars to the rest of the team.





Accessibility

Accessibility

To start, your data must be easily discoverable. No more treasure hunts or bureaucratic need-to-know barriers. To accomplish this, your data stewards—the people who most deeply understand each data set’s lineage and meaning—must build and contribute to a data catalog. This catalog will be a trusted source of truth about how your firm creates, curates, and uses data, and will help you promote its value. The more people use the data catalog, and the data it describes, the more value your firm derives from the asset.

The data catalog should include information about update cycles, schema definition, and corporate security policies. A data user, often the data scientist or application developer, should never be surprised by the data’s content, quality, or currentness. It’s best to start with an agile mindset, focusing on the basics. Building a solid foundation will be a major win and will help the team understand what’s involved. Then you can continue evolving and improving your cool new data catalog.

Data is worthless unless it’s accessible, so we need every department head to nominate one or two data stewards—the people who have the most in-depth knowledge of the data generated in your area. We’ll work with those people to create a data catalog.



Usability

Usability

Data must be usable once you find it, which means people must understand its quality and characteristics. Here again, the data catalog provides significant value. Ideally, your data catalog will include searchable metadata—descriptive information about the data set—so people can clearly understand the data’s meaning and intended use. You also need detailed descriptions of data lineage. How exactly was the data created? Ingested? How was it cleansed, transformed, or normalized? When is old data rolled off to cold storage? All this may affect

how the data is used.

Usability also includes how you present data. You don't want to be like the stone-age third-party data sites that only serve up flat-file downloads (csv, xls) or roll-ups that pre-summarize the data. Good data-driven decision-making happens when you use meaningful roles and privileges to provide current, raw data—if possible, from a data warehouse via API or direct database access—and let the data scientists do their job.

Finally, you'll need a glossary of your common business terms. For example, the team will finally have to agree on what “active user” means. Is it a registered user with an active account, or one that recently logged in? Or is it an account associated with active equipment? Without this agreement, users will claim the data is wrong. Clarity on the meaning of your business terms will clarify the meaning of your data, which will enable data-driven decisions.

We also need to consider usability. We'll start by identifying the data we want and moving those spreadsheets and tables into a proper database. We need to agree on a business glossary, too. Then, we will have to locate the right expertise to help us create useful direct data access.



Integrity

Integrity

Data integrity is all about data quality and how it's protected. Your company generates a lot of data, but it's not always curated for quality, so you'll have to clean, normalize, validate, and enhance it as it's ingested—for example, by standardizing country, province, or state abbreviations or normalizing dates to GMT.

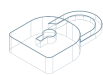
You have to ensure that users trust the data, so having a single source of truth is paramount. If customer ID fields exist in many data sets, you must choose and reference only one authoritative data set for customer IDs. The goal is complete, non-duplicative data, both within and between data sets. Users will only trust the data if they trust your approach (explained in your governance plan). Your diligence in managing this—and ongoing education, documentation, and transparency about the data—will help to earn that trust.

To better manage authoritative data, you'll need a centralized data lake. Data that is siloed and hoarded in different departments is hard to trust. Individual departments may have the deepest understanding of a data set's meaning, but that understanding should be shared in the data catalog so everyone can benefit.

Often, the final data set is sourced from several correlated data sets. Data stewards will ensure the quality and integrity of individual data sets, but a centralized data platform allows your team to curate each data set in the context of the others so it's consistent—another reason for normalizing timestamps, for instance. Without ongoing, diligent curation, your data lake will become a data swamp.

Data life cycle management also impacts integrity, as well as storage costs and overall system performance. Most warehoused data should not be kept on production servers forever; data from 10 or 20 years ago is often just clutter. Imagine the CEO's favorite dashboard with 20 years of sales data including divisions you divested five years ago. The data stewards can help determine the best horizons for each data set. You'll use that information to create a centralized long-term data life cycle policy, and then enforce it—especially for very large data sets. You can access the old data and backups from cold storage when needed.

Good decisions require good data—data integrity. The ideal solution is a centralized data lake; we'll discuss this in our next meeting. For now, consider the sources and quality of data produced in your area so we can decide on next steps. Your data stewards should be able to help.



Security



Security

Security is a key part of data governance, and the first thing most people think of, but it's only part of the picture. It does no good to lock everything down and prevent all access; you need sensible data access policies. You need to consider the value of the data as well as any applicable regulations. Do you collect personally identifiable information? Do any laws or client agreements regulate its

use? You can't assess that until you understand the data catalog and use restrictions.

Then, you have to ensure that your data platform can apply highly specific role-based security, independently of the applications that access the data, and that it can secure data at both the row and column level. You may even need security-level access for parts of your data catalog, such as metadata or schemas. If the complete schema might reveal a column for something that is clearance-only, the catalog's metadata must reflect the user's access.

To secure your new data lake, you'll have to add your data stewards to your governance team. They will describe in detail the roles, privileges, and restrictions needed for each data set, and create a system to capture any changes over time.

You also need a data privacy officer—someone who can be the internal expert on applicable laws such as General Data Protection Regulation (GDPR) or the California Consumer Privacy Act (CCPA). There may also be regional regulations to consider.

Finally, you need to monitor all access for audit purposes. Access logs are an important part of any audit. That means judicious logging of all incidents and making log data easy to retrieve—both for audit purposes and for notification of any data breaches. No cutting corners on the protection of customer data, especially things like credit card information, social security numbers, or health records. You don't want your data set to end up in the news.

Finally, we need to secure our data lake and ensure we are in compliance with our contracts and any governmental regulations. Once we understand what we're protecting and the applicable requirements, we'll need a data privacy team to help us stay compliant.

Getting to Work

Until now, there has been no common language for data across your organization. Data hides in departmental silos, and the usual response to a data request is "Tell me what report you need, and I'll put in a service order."

Forehead, meet palm. Your CEO deserves better. Your company deserves better.

Despite the simplicity of the CEO's request, you know this isn't about dashboards; it's about making better decisions. The CEO wants not only the executives but the entire team to be more informed and to have access to consistent, trusted data. Without carefully planned and consistently executed data governance, that won't be possible; people will get bogged down in endless questions about where the data came from, what everybody means when they say "revenue," and whether or not that spreadsheet is useful. Most importantly, if the data is wrong—and it often is unless you deeply understand it—the decisions will be wrong.

You have a tremendous opportunity here. The CEO probably expects you to do exactly what was asked—grab a bunch of data from disparate sources and make a few flashy charts. But you are in a position to make a strategic contribution by implementing a solid data governance plan and changing your corporate culture when it comes to data.

Time to make some more phone calls.

About productOps

productOps is a consulting firm that delivers its strategic insight as a software 'product', and has been designing, building, and implementing data platforms since before anyone thought to call them that. You can reach us at dataqueries@productops.com.

Contact us



Ryan French
831-214-4714
ryan.french@productops.com



Bob Cagle
831-466-3000
bob@productops.com