Extracted from:

Designing Data Governance from the Ground Up

Six Steps to Build a Data-Driven Culture

The Pragmatic Bookshelf

Raleigh, North Carolina

The Pragmatic Programmers

Pragmatic exPress

Designing Data Governance from the Ground Up

Six Steps to Build a Data-Driven Culture

Lauren Maffeo

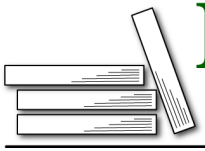
Edited by Brian P. Hogan

Designing Data Governance from the Ground Up

Six Steps to Build a Data-Driven Culture

Lauren Maffeo

The Pragmatic Bookshelf
Raleigh, North Carolina



Pragmatic Bookshelf

For our complete catalog of hands-on, practical, and Pragmatic content for software developers, please visit *https://pragprog.com*.

The team that produced this book includes:

CEO: Dave Rankin
COO: Janet Furlow
Managing Editor: Tammy Coron
Development Editor: Brian P. Hogan
Copy Editor: Karen Galle
Layout: Gilson Graphics
Founders: Andy Hunt and Dave Thomas

For sales, volume licensing, and support, please contact *support@pragprog.com*.

For international rights, please contact *rights@pragprog.com*.

ISBN-13: 978-1-68050-980-9
Encoded using the finest acid-free high-entropy binary digits.
Book version: P1.0—January 2023

What Are Data Stewards?

Data stewards own the strategic and tactical decisions for data within their respective business domains. Stewards serve as trusted advisors for their data, providing key context and nuance as needed. For example, a director of customer success could serve as a data steward by owning and offering key insights about customer data that your business prioritizes, like average response time and the number of new accounts per month.

You might wonder if data stewards must hold technical roles. After all, data scientists, engineers, and similar roles work with all types of data each day. Aren't they most qualified to serve as stewards? The answer is sometimes, but not always.

Data stewards bring their own unique value to your business. This includes deep knowledge of the company strategy, customer pain points, collaboration processes, and cultural nuances. By working in a diverse range of roles, they lead areas of the business that your data team might touch tangentially, if at all. Perhaps most crucially, data stewards are experts in their unique domains, whether that be sales, marketing, product, or legal. This knowledge is invaluable to colleagues in data engineering, data science, and data architecture; they need those insights to support their own work.

For example, let's say your data team must re-design the taxonomy within a database about cars. Before that content reaches its database, a team of statisticians collects, downloads, cleans, and analyzes the data for their specific products. This expertise varies per statistician. Whereas one statistician owns the data for BMWs, another owns the data for Audis.

Database users tell your user research team that it's too hard for them to find the data they need within this database. To help users get more value from your database, your data team must redesign the taxonomies for BMW and Audi products. Before your data team can start this crucial work, they'll need several questions answered:

- How is the data for these commodities collected?
- Who collects this data, and where is it stored before it reaches the user database?
- How many databases does this data live in? Are they on-premise or cloud databases?
- How often is new commodity data released to the public?

- How are these car-specific data releases prepared, and who prepares them?
- Is this preparation process automated, manual, or both?
- Which quality assurance checks do statisticians use, and which colleagues participate in these processes?

To find accurate answers to these questions, your team will rely on statisticians who can steward their respective data. (In this case, data about Audi and BMW vehicles.) These stewards will work with your data team to help them learn the answers. Data stewards could help your data team do the following tasks:

Define attributes

Data attributes (also known as metadata) define properties of objects, elements, or files. Data stewards are often the best people to write, manage, and share attributes for the data in their domains. For example, a customer service director might manage details such as customer IDs, addresses, and purchase histories. If this director serves as a customer data steward, they can keep clean, organized, secure records of this data to share with colleagues as needed.

Build data models

If data stewards lack a data architect's expertise, they can still help architects with their work. For example, a data architect who's redesigning a database will need to understand the IT systems that this database must integrate with. A steward who manages these systems can give crucial context to help the architect build better architecture that supports the right integrations, taxonomies, and so on.

Create data dictionaries

Think of data dictionaries as centralized spots for information about data. A great data dictionary helps a system's users find all the details about data that they need in one place. Data stewards can play a key role contributing to these dictionaries as writers, editors, and advisors.

Not all data stewards serve the same function. Where possible, it helps to categorize data stewards as serving either business or technical roles. Separating data stewards into business vs. technical stewards might sound intuitive, but it's often a big roadblock on organizations' data governance journeys.

No matter how nimble a business thinks it is, the truth is that all but the smallest teams have more silos than they think. This leaves most teams with little to no knowledge of how their colleagues in other departments work, or

what they work on. And nothing reveals those gaps quite like searching for the right subject matter experts to fill decision-making roles about data.

This is also a hard task because many data stewards lack the autonomy to make organizational change. Despite their status as experts of the data in their domains, they often lack the power to make choices that affect the business strategy. At best, they've built strong relationships with those who have organizational sway, and they can effectively persuade these colleagues to consider their advice. At worst, their advice languishes. To avoid this, a big part of your job will involve giving your stewards the visibility and recognition they deserve.

You'll get some tips on how to do that soon. For now, here's a quick overview of how business data stewardship roles differ from technical roles:

Business stewards own data for business processes and workflows. They own the data classification scheme for data and metadata which falls under their areas of expertise. For example, in our car database, a statistician who produces data releases for BMWs would serve as a steward for BMWs' taxonomic hierarchies. This involves these tasks:

- Collaborating with colleagues to make sure that these hierarchies are accurate, organized, informative, and intuitive
- Collecting data and confirming that this data/metadata meets organizational quality standards
- Reviewing and assessing current taxonomies for commodity line items
- Writing the documentation required for these commodity line items
- Leading data governance trainings for their departments
- Defining, developing, and conducting data quality assessments against pre-defined business metrics

Technical stewards own data for systems, pipelines, and implementation. These stewards help colleagues learn how to model, create, maintain, and transfer data between systems. Technical data stewardship tasks might be these:

- Maintain all requirements for system implementations.
- Manage data marts, data warehouses, data lakes, and data mesh.
- Build and integrate the systems mentioned previously.
- Define parameters for metadata.
- Confirm which data they need to assign different types of content or assets.

- Set limits for which data users and admins can work with.
- Move data from an on-premise to a cloud-based environment, or from one cloud environment to another.
- Build data models and algorithms that achieve business priorities.

Within these broad business and technical categories, data stewardship roles get even more specific. As you read the role descriptions that follow, consider them all within the context of the data framework you referenced in Chapter 1, Find Your Data Framework, on page ?. Your data stewardship team will fulfill this framework by owning data in specific business and technical domains.

All lines of business which produce and/or manage unique data should have designated data stewards. Sales, marketing, and customer success are three teams which manage data that keeps your business running. You should thus have at least one senior member of each team serving as a data steward who can own the data in their respective business domains.

Your business might need more or fewer stewards than this list, which is fine; adapt your stewards to your own business needs and restrictions. The key takeaway is that any stewardship role should align with at least one aspect of the seven-part data framework. With that in mind, here's a non-exhaustive list of roles that data stewards can serve:

Council Chair Data Steward

Your data stewardship council chair plays its key leadership role. The chair leads all meetings with stewards and represents the data governance council at meetings with folks throughout the business. The council chair approves requests to start projects or buy new data tools, leads meetings with the data governance council about its direction, and owns the data governance strategy. At a high level, they perform the following tasks:

- Build and implement a strategic data management plan for data governance council members to execute.
- Oversee all projects and initiatives related to organizational data.
- Assign responsibilities to data stewards for respective technical and business roles.
- Share communication about changes to data initiatives.
- Represent the data governance council at C-suite and cross-departmental meetings.

- Explain the organization's response to security breaches.
- Define what "quality data" means in your organization, including which data sources to use and avoid.

Security Data Steward

Security data stewards set data usage and security policies in partnership with the data governance council chair and C-suite. Security data stewards liaise between the business and technical sides of your organization to set and share security updates throughout the broader business.

Security stewards perform the following tasks:

- Oversee all security requirements to safeguard data.
- Assign appropriate security classifications to all company data based on sensitivity.
- Collect all security-related data artifacts for legacy and new projects.
- Create security artifacts as needed.
- Assess current security requirements, and adherence to these requirements.
- Make security findings available on request.
- Define data lockup requirements.
- Conduct gap and impact analyses of all new security requirements.
- Compare security requirements against current and proposed data environments.
- Audit and approve all proposed production changes against security risks.

Your security data steward should review any new project or tool that your council approves. They'll assess each request against the organization's security standards, offering guidance early to avoid technical debt.

Ethics and Transparency Data Steward

This steward ensures your data governance policies meet the clear, defensible, and documented standards discussed for data use in Transparency and Ethics, on page ?. Your ethics and transparency data steward works directly with the data team to help data analysts, scientists, architects, and engineers consider how their work might impact customers in unintended ways.

The ethics and transparency data steward might do hands-on work with data if they have a data science or engineering background, but this isn't required to drive change in the role. Your ethics and transparency data steward will need to consider a broad range of digital ethics risks per data project, then share those risks with the data governance council. An ethics and transparency data steward should perform these tasks:

- Write compliance standards for how your organization stores, shares, collects, and protects data.
- Map possible ethics risks per data project back to your organization's business and data mission statements.
- Document decisions for how to manage data, including rationale for how these decisions serve the business and customers alike. (This is especially important, as customer well-being often stands at odds with what's "best" for the business.)
- Train employees across the organization to practice data ethics, from C-suite leaders to individual contributors.
- Ensure that your organization's data ethics practice meets broader legal and legislative compliance standards.
- Confirm that all technology your business buys and uses meets your organization's own compliance standards.

Many ethics and transparency stewards have backgrounds in law and/or compliance, which helps them assess risk and predict which data use might breach civil rights.

Documentation Data Steward

Your documentation data steward is your council's resident writer. They manage the council's documentation repository, which stores everything from council meeting notes to business requirements for user acceptance testing per project. Within this stewardship role, they perform these tasks:

- Write consistent data definitions.
- Train colleagues across the organization on best practices to document data-based work.
- Store this documentation in the appropriate area(s).
- Write, edit, and manage documentation for all data-based decisions and initiatives.

- Manage artifacts like data architecture diagrams, data model(s), and data dictionaries.
- Edit data artifacts as needed in partnership with technical stewards.
- Organize all high-level data assets, such as data dictionaries and metadata catalogs.

This data steward might not write all the documentation themselves; part of a data governance council's value is having folks across the business own data-specific work so it doesn't fall to one person. In cases where they don't write documentation themselves, the documentation data steward still owns the repository and stores documentation appropriately.

Compliance Data Steward

Compliance data stewards often bring legal backgrounds to their data governance work. This steward tracks regulatory changes involving data and ensures data usage meets those changes. They work closely with all roles on the data stewardship council, from writing compliance standards for the documentation data steward to reviewing data transparency requirements with your ethics data steward. The compliance data steward performs these tasks:

- Reviews regulatory frameworks for data across markets where your business serves users.
- Interprets diverse laws to confirm data-specific rights for businesses and consumers.
- Ensures that your business follows essential laws for data management per project and vendor selection.
- Communicates with data stewards and the data team to ensure that everyone considers compliance pre-resource allocation.
- Reviews models during the data use life cycle to ensure that the data team follows data quality standards before, during, and after production.

If some of these titles sound familiar, it's because key stewardship roles align with the data governance framework you explored in Find Your Data Governance Framework, on page ?. When building your own data stewardship team, make sure you have alignment between stewardship roles and the respective parts of your data framework. Assigning colleagues to co-own your framework will help bring it to life.

Depending on your organization's size, you might not have enough stewards to fill roles that address all aspects of the framework. In those cases, look for

opportunities where one steward has the expertise to own two parts of the framework. For instance, there's a lot of natural overlap between the ethics and transparency data steward and the compliance data steward. If you have an attorney on staff to serve as the compliance steward, ethics and transparency can fall to them as well.

When trying to fill roles per part of the framework, consider who's already doing this work. A data steward who leads customer success could be the ideal fit for your education and training data steward role. This empowers them to help all colleagues in your organization learn what data governance is, why it's relevant to everyone's roles, and how to ask data stewards for help. Customer success teams spend their days speaking to clients, triaging questions, and showing diverse people how to use your products. Education and training is already a key aspect of their work. This aptitude puts them in a natural spot to steward education and training for data governance.

Likewise, a marketing leader can serve as your collaboration and culture data steward. Marketers explore, create, and lead opportunities to find the right audience for your products. The best marketers are natural storytellers who can seamlessly share how their products meet your needs. You will need that storytelling prowess to share why data governance is a team sport, and guide your company culture through the change required to make it work long-term.

You will know your data stewardship council is on the right track when each aspect of the seven-part framework is accounted for. But if you're still in the early stages, you might wonder how to hire data stewards at all. The great news is that you don't have to.