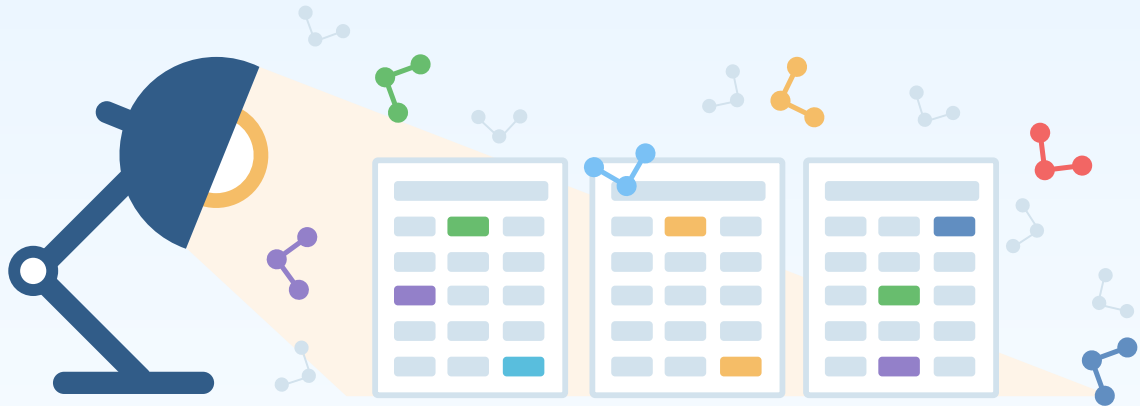


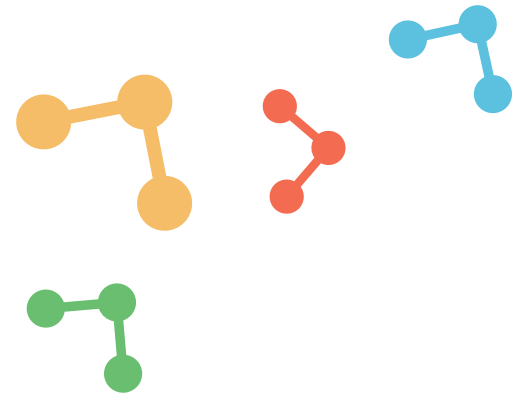


data.world



How to Plan and Launch Your Modern Data Catalog





Let's get started.

BUILD A STRONG FOUNDATION

Implementing a data catalog helps every member of your data community discover and use the best data and analytics resources for their projects, achieve faster results, and make better decisions. Data catalogs illuminate tribal knowledge and spur collaboration, both of which are key elements of [Collective Data Empowerment](#). Catalogs are an essential foundation to building your data-driven culture.

It's important to invest time in planning your data catalog implementation. A strong foundation will support your entire organization as it grows and your data needs become more complex.

ABOUT THIS RESOURCE

This guide helps you determine who should help, which resources you'll need, and what success looks like for a catalog launch.

We'll walk you through the basic steps to get your data catalog up and running, discussing the challenges you might face along the way and how to address them. We'll also provide metrics to measure its impact, and highlight strategies to engage your data stakeholders so they value and adopt the data catalog in their work.



Hint: This worksheet is an editable PDF. Type in the text fields and check the boxes as you go. You also have plenty of writing room if you prefer to print it out and work with pen and paper.

Step 1: CHOOSE A PILOT PROJECT

Avoid the urge to immediately onboard your entire organization. Instead, begin with a clear, well-defined analytics pilot project. This way, everyone has a common goal while you get started, and other teams can learn from your best practices.

WHY THIS WORKS

- **Reduce risk:**
Start with a clearly scoped and relatively quick project to learn what works and what doesn't before any challenges affect the entire business.
- **Learn by doing:**
Create a blueprint that other teams can use as you expand data catalog usage. You'll be able to share measurable results, processes, and recommendations with the next team or project. This makes each stage more efficient.
- **Build buzz:**
When you're ready to onboard other teams and projects, they'll already be familiar with the value of a catalog for data and analysis. Essentially, you're creating a miniature case study with your pilot project's success.

WHO TO INVOLVE

Put together a cross-functional selection team to start building awareness and ensure that the best project candidates are considered. Include sponsors from the Chief Data Officer's organization, data-rich lines of business, and IT.

Your selection team should be composed of people who know your business best, whether in an executive position or on the front lines. In some cases, this is the same team that then works through the pilot project itself. Looping in these sponsors from key areas of the business right from the start will spur buy-in, making it easier for you to get initial resources and build internal buzz.

COMPLETE THESE TASKS BEFORE MOVING ON TO YOUR NEXT STEP:

Identify the members of your selection team:

Confirm that you and your selection team have agreed on a project that satisfies the following criteria:

High priority for the business to solve right now (vs. later)

Clear scope that is not overly complicated

Use case that will resonate with other teams

Step 2: ENGAGE THE RIGHT PEOPLE

Identify the experts who will complete the actual tasks of creating, managing, and using your data catalog. Again, they may be the same team that chose the project, or they might be representatives from the same team. Every business is different.

WHY THIS WORKS

This will be the first team trained to use the data catalog, and their feedback is critical to fine-tune your broad implementation strategy later on. They will also be the evangelists and subject matter experts for future catalog use across your organization.

WHO TO INVOLVE

The pilot team should have participants from the CDO's team (or the equivalent data-oriented leader in your business), from IT, and from project-relevant line of business areas.

Again, there may be some overlap with your selection committee from the previous step. Ideally, this team should involve people that will plan and conduct the day-to-day work on your pilot project. Consider the **data engineers** who will set up your data sources, **data stewards** who contribute governance policies, and the **business analysts** and **data scientists** who add business context to the data in the catalog and use it for analysis.

USE THE CHART ON THE FOLLOWING PAGE TO IDENTIFY YOUR PROJECT'S STAKEHOLDERS.



ENGAGE THE RIGHT PEOPLE

ROLE	CONTRIBUTION	NAME(S)
Data Manager	<i>Curates and manages access to the data, or knows where to find it. (For example, data stewards or data engineers.)</i>	
Subject Matter Expert	<i>Not necessarily technical, but understands the market, product, or topic involved.</i>	
Data Practitioner	<i>Answers the question using data and related context. (For example, data scientists or business analysts.)</i>	
Data Consumer	<i>Makes decisions from the analysis produced; often asks a question that kicks off a data project.</i>	

COMPLETE THESE TASKS BEFORE MOVING ON TO YOUR NEXT STEP:

Brief the people needed for each step of your pilot project.

Make sure they understand [what a data catalog does](#) and how it can be used with their existing tools to accomplish their goals.

Confirm your stakeholders have committed time to making this pilot a success: it's a specific priority for them and they have manager approval where needed.

Step 3: SELECT AND CONNECT DATA SOURCES

Populate the data catalog with the data sources you need to achieve your pilot project's objectives. As before, resist the urge to connect every single data stream in your company: focus on the sources you need to get the project done now. You can always add more later.

WHY THIS WORKS

- **Data is brought into the light:**
Your catalog automates the collection of data and associated context (e.g., who owns it, how it's been used, what it means). This step brings together all the data your project needs to be successful, no matter where it lives.
- **Data is documented and trustworthy:**
Data managers and consumers can easily validate each resource's quality and lineage and provide associated context for others to use data effectively.
- **Data becomes instantly more useful:**
Reconcile copies of data into single, reusable sources of truth. When your data is added to a catalog, it turns into a series of building blocks that can be assembled to answer one question after another, all from the same reliable source.

WHO TO INVOLVE

Everybody on the pilot team should actively participate in setting up data sources. Technical team members populate the data catalog from a variety of data sources such as applications, databases, BI tools, logs, and files. Business users provide validation and business context for the sources and also contribute existing analytical assets such as queries, dashboards, and machine learning models as additional sources. Together, you're creating a bank of trustworthy, validated, context-rich resources.



Hint: Don't worry too much about file types. A good catalog should automatically normalize data as it's ingested so you can immediately run queries across data files in multiple formats without extra prep work.



Hint: Use [The Enterprise Data Tech Stack Audit](#) to discover any potentially useful data sources already being used. Note any additional sources you discover along the way and may want to add in later.

COMPLETE THESE TASKS BEFORE MOVING ON TO YOUR NEXT STEP:

Identify and add a few (1-3) data sources needed for pilot project success.

Confirm that your team members can all access the data they'll need for the project.

Configure your catalog to automatically track and sync changes in data sources so your team always works with the freshest data.

Annotate your data so it's easy to understand and use:

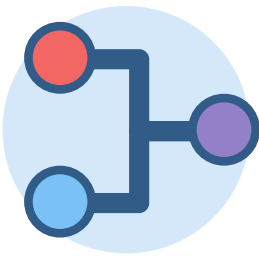
- Add descriptions: where does this data come from, and what's in each column?
- Note key stakeholders: who owns the data and who should field questions about it?

Step 4: EDUCATE AND DRIVE USAGE

Once you have identified the project goal, engaged the pilot team, and added data to the catalog, your goal will be to help the pilot team use it as part of their daily workflows and toolchains.

This means showing them how the catalog fits into and improves their processes, increasing efficiency and data quality overall. Then, stay engaged: answer questions and make suggestions. Follow up and hold everyone accountable for their contributions, whether that's data work or providing feedback. The pilot will be most successful when everyone contributes.

The guides below help you answer common questions from stakeholders on your pilot team. (Getting questions you don't quite know how to answer? [Let us know](#) and we'll help.)



Data Manager

Common Roles:

Representatives from the CDO's team, data steward, data quality manager

Typical Responsibilities:

Enable cross-organizational data access, visibility, and reusability in accordance with governance policies

Common Questions:

Q: How do I adapt our information governance policies to incorporate the different types of data sources that will now be visible and consumable by users of the data catalog?

A: As part of the implementation, you should engage with data owners across lines of business (sales, marketing, finance, legal, HR, etc.) and IT to ensure a shared understanding of the goals of the data catalog and create a plan for data access SLAs and other governance policies.

Q: How will others determine if data is trustworthy and useful?

A: To help validate the data, catalogs document feedback from peers on each dataset's accuracy and usefulness, and often include quality metrics determined by your governance team, too. You can also see how others have used data in similar projects, replicating or reproducing their analysis to move other projects forward. And because data is synced from its original source, you'll know it's up to date.



IT & Engineering Leader

Common Roles:

Representatives from CIO or CTO teams such as data engineer, data architect, BI developer

Typical Responsibilities:

Bringing different data sources into the catalog and ensuring that the data in the catalog remains up to date

Common Questions:

Q: How does this work alongside my existing data repositories? Will I have to learn or adopt a new set of tools to connect data to the data catalog?

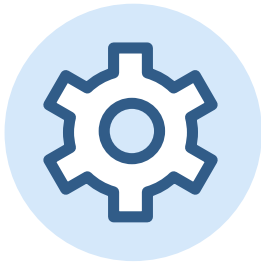
A: Most data catalogs come with pre-built connectors to ingest metadata from the most common data sources. This allows you to continue working with the data in place while also making it easier for others to discover, collaborate over, and contribute value to the data over time. The best data catalogs also allow you to ingest and work with data in the platform. Some data catalogs may also work seamlessly with your existing integration tools and sources.

Q: How does the metadata stay fresh for all the data sources?

A: Along with connectors for initial loads, the best data catalogs include automated scanners that can be configured to update changes with data sources.

Q: How are new data sources added to the catalog once the initial pilot project is started?

A: If you discover additional sources you'd like to include in the catalog, most teams will work with an organization's data steward or similar role to curate the data. However, the best data catalogs make it easy for even non-technical roles to contribute their data resources.



Line of Business Experts

Common Roles:

Data scientists, business analysts, and subject matter experts from specific areas in the business such as Sales, Marketing, HR, etc.

Typical Responsibilities:

Use data to quickly make decisions that power day-to-day business, supplying critical data and expertise on the business, market, and customers

Common Questions:

Q: How does this work with the tools you already use?

A: The most comprehensive data catalogs offer integrations with tools across the entire data lifecycle so everyone can contribute value, regardless of role, technical ability, or the tools they use. This lets you seamlessly discover and validate data sources, as well as collaborate with experts without leaving your existing tools.

For example, let's say you're working on a new marketing analytics project to build a customer demand forecasting model. Your tools of choice are Excel, Tableau and SQL queries. But before you can use any of those, you need to pull together data from many sources that you know about: historical sales database, CRM, web analytics, and log data. A catalog helps all your data sources speak the same language, making them more interoperable. You'll save time and be able to immediately dive into the data you've found. Then, you can share your analysis back to the catalog to help others.

Q: How do you find out about other potentially helpful sources used by other teams, or discover additional datasets that you should be using (e.g. data from social media and partners)?

A: Augment your analysis (or springboard off the work someone's already done) with data contributed by other teams in the organization. Find approved data sources in the catalog and filter data by business context so you can instantly tell if it's relevant.

Q: How do you determine if the data you are using is trustworthy?

A: To help you validate the data, catalogs document feedback from peers on each dataset's accuracy and usefulness and often quality metrics determined by your governance team.. You can also see how others have used the data in similar projects, replicating or reproducing their analysis to move your own projects forward. And because data is synced from its original source, you'll know it's up to date.

Step 5: MEASURE SUCCESS

Work with your pilot team and executive sponsors to identify and track key metrics to quantify the impact of using the data catalog. (Actually, it's not a bad idea to do this before the other steps in this launch guide. We recommend using the [Modern Data Project Checklist](#) because it helps you determine success metrics and objectives up front.)

WHY THIS WORKS

- **Understand unique viewpoints:**
In addition to quantitative metrics, you should also capture anecdotal feedback from each pilot team member, both pros and cons, on how the catalog helped them achieve their project objectives. [The Modern Data Project Retrospective](#) provides outlines and exercises to help.
- **Refine your process before onboarding others:**
These metrics and anecdotes will be key in providing feedback on the efficacy of the catalog implementation and enable you to make improvements. Essentially, you're working out the kinks with the pilot team before bringing in others.
- **Provide proof:**
Quantifiable results showing ROI help you market the data catalog for broader adoption throughout your company.

WHO TO INVOLVE

Your pilot team. Simple as that. Not every team member will have specific metrics to measure, but everyone should be able to share how the project impacted their work.



Step 5: MEASURE SUCCESS

COMPLETE THESE TASKS BEFORE WRAPPING UP:

Tracking the catalog’s impact on team productivity, organizational culture, and overall business results will be an ongoing feedback loop with continuous tinkering. There are different types of metrics that can be tracked; a few examples are listed below. Some of these, such as the Productivity metrics, should be tracked before and after using the data catalog, if possible. Some of the metrics, such as Usage, may only become trackable once you have a data catalog in place.

Productivity:

- Average time to complete data selection phase of analytics project
- Percentage increase in reuse of datasets and analytics assets
- Percentage of datasets and analytics assets with quality ratings

Data-driven Culture:

- Increase in number of data consumers throughout organization
- Percentage of data consumers that contribute to data curation

Usage:

- Most popular data sources
- Lineage of each data source
- Top users of different data sources and analyses

Business:

- Impact on revenue due to faster time to insight
- Lower cost to find and validate data with automation and self-service

TRACK YOUR IMPACT

METRIC	BENCHMARK (before catalog)	RESULTS (after pilot project)
Productivity		
Data-driven Culture		
Usage		
Business		

Next Steps:

As a result of this framework, you should have a solid example of how your business can plan, build, deploy, manage, and grow your data catalog:

1. *When you chose a pilot project, you identified a sample use case that explains how a data catalog can be used by others, connecting it to actual business outcomes.*
2. *When you engaged the right people, you created a team of experts that can continue to support, educate, and evangelize the value a catalog brings to each part of the business so it's easier for others to understand and integrate the catalog into their daily workflow.*
3. *When you selected and connected data sources, you created a central, trustworthy resource for essential data that others can immediately access, use, and collaborate over.*
4. *When you focused on educating and driving usage, you proactively addressed common questions, creating a bank of information other teams can leverage as they begin using and contributing to your catalog.*
5. *When you measured success, you validated the benefits of more people collaborating over data together. Your results quantify the business impact that comes from more inclusive, connected data and analysis.*



Try these tactics to continue successfully launching your data catalog across the business. If you have other questions along the way, [data.world](#) can help.

- **Share your success.**
Don't hold back! Educate others on the value of creating a central, collaborative, and inclusive resource for data and analysis. Make sure to explain the business problem you solved (Step 1), who helped accomplish that (Step 2), and how you know it worked (Step 5).
- **Stay one step ahead.**
Have your next project already in mind. Can you springboard from the project you just completed, using the data to answer a different question? Are there upcoming campaigns, events, or deliverables you could support? Above all, be ready to keep moving. Keep the excitement going!
- **Stay connected with your pilot team.**
As leaders, you all provide valuable cross-functional visibility, bringing forward questions, suggestions, and requests from your teams. The best catalogs bring together data, people, and analysis; the pilot team plays an important role in keeping all three aligned.



Ready for more Collective Data Empowerment?

This tool is most effective when used in conjunction with other tools, practices, and strategies to make everyone you work with more productive with data.

[Get your comprehensive field guide to Collective Data Empowerment from data.world.](#)

About us

data.world is the modern catalog for data and analysis. It wakes up the hidden data workforce by replacing outdated barriers and disparate silos with deep connections between data, people, and analysis. When every employee can find, understand, and use data, your advanced practitioners can focus on high-impact work, and you can build a data-driven culture, faster. data.world is a Public Benefit Corporation headquartered in Austin, Texas. Visit [data.world](#) for more information and expert guidance.



data.world