Informatica

Six Strategic Steps to Democratizing Data

A Chief Data Officer's checklist for enabling data-driven decisions

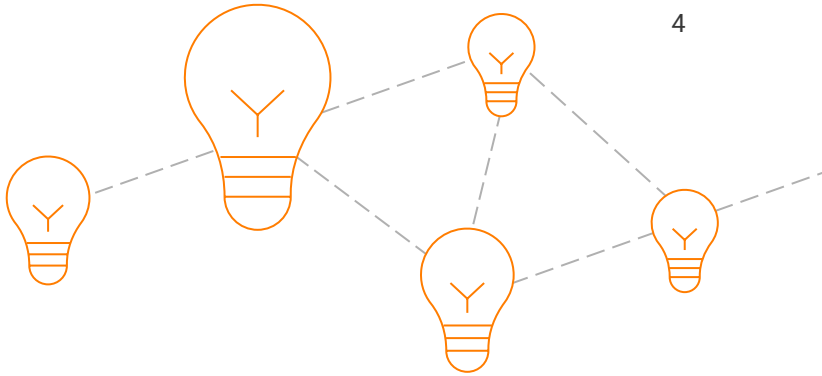
Contents

Tip: click to jump straight to any section.

Introduction

Driving Data Empowerment

Driving Data Empowerment



Your organization has tremendous opportunities within its grasp: opportunities that spring from connecting your trusted data to your people so they can use it to accelerate your value creation agendas. Intelligence gleaned from this data can be used to power innovation that in turn drives greater customer loyalty, more accurate analyses, extra-competitive product and service offerings, dramatically more efficient operations, and more.

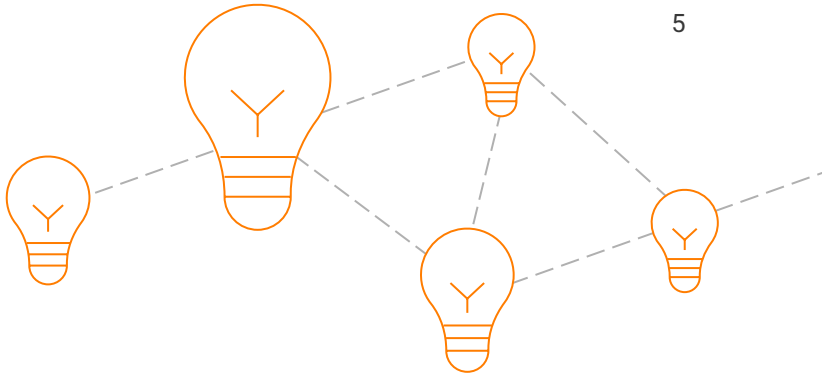
- Address volatility and accelerate digital transformation with an increased velocity of change.
- Support your business with faster, broader, and easier data sharing and use, with real-time aggregation and syndication of data to support operational reporting and analysis.
- Empower more data-driven decision-making by your front-line workers by enabling scenario modeling and analysis—with broader context—using new data sources such as third-party and data from other business divisions.
- Achieve better scenario planning with access to meaningful and quality data that is currently created and stored in business units, geographies, and organizational functions outside of your organization's systems
- Incorporate greater use of rolling forecasts in 2021 and beyond, and move more strategically with daily and weekly forecasting and planning cycles.

Now is the time to prepare. Now is the time to democratize your data.

But what does that mean, "to democratize data?" Instead of keeping data locked up so only data scientists and other experts can access it, you make it available to the masses. You empower all the knowledge workers in your organization to use it to perform their jobs with the transparency needed that aligns to your organizational policies for data use, and to derive value from it, responsibly.

On a practical level, this means creating **a data marketplace.** As we shared in The Four Big-Time Benefits of Building a Data Marketplace, this will deliver untold competitive advantages to your organization. The opportunities—and the volumes of data available—will be truly endless.

Driving Data Empowerment



Data, Data, Everywhere

According to IDC, we're seeing a 61 percent compound annual growth rate (CAGR) in worldwide data at this moment—a rate of increase that will result in 175 zettabytes of data worldwide by 2025. To express this in down-to-earth terms, a single exabyte alone could stream the entire Netflix catalogue more than 3,000 times. Now, a zettabyte is approximately 1,000 exabytes. So that's every film and TV show on Netflix played more than one million times. And then you multiply that by 175.

If you're not stunned by these numbers, consider that by 2025, this will equate to 40 times the number of stars in the observable universe, according to the World Economic Forum—and furthermore, says Forbes, the volume of data at that point will double every twelve hours.

But this also raises challenges as you pursue the very real value that your data offers. Specifically, **data governance** challenges.

Without proper controls and policies in place to protect and manage these vast and growing volumes of data, its quality and trust will be compromised. And once the quality of your data is no longer trusted, your business—now more dependent on data-fueled digital processes than ever before—could flounder.

Put another way: before you can build your data marketplace and truly democratize data, you need a way to ensure that your data is trusted so that your results can, in turn, also be trusted.

In this workbook, we will outline six strategic steps to help you achieve the confidence you need to do this:

Step 1: Document and collaborate
Step 2: Discover and curate
Step 3: Cleanse and master
Step 4: Protect and monitor
Step 5: Provision and consume
Step 6: Set up shop and deliver value

Now, we admit that these six strategic steps will require some effort. It will take some coordination of stakeholders, and remediating issues that call for responsible data stewardship. You'll need the right tools to intelligently automating key processes. ***But,*** if you can master these steps, you will truly democratize data. You will build a thriving data marketplace, and accelerate data-driven decision-making throughout your organization. And not incidentally, by doing all this, you can also reap the benefit of establishing the Chief Data Officer and his or her team as leaders who deliver measurable competitive value to the business.

Six Strategic Steps

Step 1. Document and Collaborate

1

The first step in democratizing data is to begin by establishing a foundation for data governance and documenting it. With that done, you then need to nurture a collaborative data culture that will build on that documented foundation. That's because even the most carefully thought out governance plan won't work without the right collaborative culture.

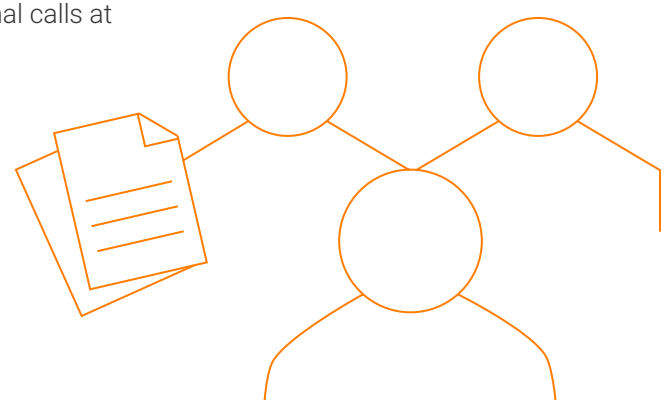
One way to establish the right collaborative culture is to create a data governance charter. This begins by asking a lot of critically important questions, such as: What does the data mean? Who owns the data? What is it good for? What policies govern it? Who will use it, and how? The answers to these questions, allow you to create detailed data policies and ensure that your team has the right processes in place to support those policies. And of course, document everything for alignment and transparency: the policies, processes, business glossary, stakeholders, and metadata.

But asking the right questions and writing down all of the answers is only half the battle. If the work is done in a vacuum, it's less likely to be followed. According to McKinsey, "When people are excited and committed to the vision of data enablement, they're more likely to help ensure that data is high quality and safe."

And this is where a **data-driven culture** comes in. Yes, it's trendy, but it happens to be key to today's competitiveness: Data is what you're going to cite in high-stakes meetings to make salient points. Data has become the deciding factor when employees make decisions of all kinds, not just big, strategic ones. And a data-driven culture will reward workers for using data responsibly, instead of allowing the HIPPO (highest-paid person in the organization) to make final calls at odds with what the data says.

The cornerstone to fostering a data-driven culture? Collaboration. When the user community and IT leaders work on data governance together aligned to a shared set of goals, both sides win—and the governance program pays substantial dividends.

Unfortunately, most organizations haven't completed this first, very basic step. The Boston Consulting Group (BCG) found that 60 percent of businesses rated their data governance capabilities as "underdeveloped." And much like erecting a skyscraper, if you don't start with a foundation you can trust, your top floors—and your goal of democratizing data—will come crashing down.



Step 1. Document and Collaborate

1

Deutsche Leasing: Collaboration in Action

Collaboration between business and IT teams was the cornerstone of data governance success for Deutsche Leasing. The need was urgent: with more than 20 financial systems around the world, each with its own data model adjusted to local legislation, the firm found itself facing a seemingly impossible task: deciphering which version of "truth" was correct. No matter how good the IT experts involved were, only the savviest business users were able to come up with answers. So, reasoned Moritz Schlee, a senior IT project manager and Alexander Khrebtishchev, head of global reporting for Deutsche Leasing, why not start their data-driven journey from the business side?

Deutsche Leasing took a "stitch-through" approach to data governance that consisted of first nominating business experts—called business data owners, or BDOs—from each department, and asking them to write down clear definitions of the data, along with the key performance indicators (KPIs) that were broadly used within their domains. Then the IT department agreed to look up and deliver data to users—but only when the definitions and a designated BDO were in place. Deutsche Leasing used Informatica Axon Data Governance to present the defined business attributes together with automatically linked technical fields and delivered data to users via the Enterprise Data Catalog. Using Axon and then the Catalog allowed business users to look up specific terms they were familiar with, which would intelligently connect them to the right data in the IT environment—rather than the other way around. This business–IT collaboration proved essential for Deutsche Leasing to begin treating data like the valuable asset it was.

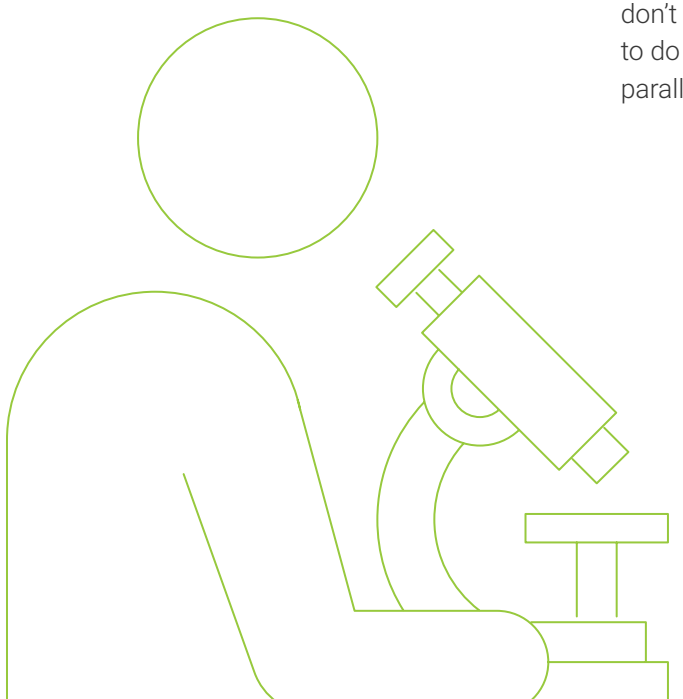
Checklist

- [ ] Do you have a data governance program today? (No? Then check out The Data Governance Program Workbook)
- [ ] Do you currently leverage a data stewardship tool to document policies, processes, glossaries, and stakeholders?
- [ ] Is the tool automated and connected to the rest of your enterprise?
- [ ] Would you call your culture data-driven? If it isn't, why do you think that's the case?
- [ ] Are your data professionals and your data users collaborating on policies and practices?

Step 2. Discover and Curate

2

Once you have your governance framework in place and documented, and a data-friendly, collaborative organizational culture that encourages users to follow it, it's time to actually find the data.



NOTE: We acknowledge that you could flip Step 2 and Step 1, as sometimes data professionals will choose to locate their data first, before establishing policies, owners, and more. The order that's right for your organization typically depends on how well your data leaders understand your organization's data landscape. If you have a good handle on it, then you should follow the steps in their current order. If you don't have that understanding, it may be wise to do Step 2 first, followed closely—or even in parallel—with Step 1.

For those who choose to discover your data first, you're not alone. Indeed, we frequently see that not being able to find their data is a major thorn in many businesses' sides. Many have no idea what data they possess, whether it should be governed as "business critical," or where it resides. This impacts the company directly, because if business professionals can't immediately find the valuable, trusted data they need, they will take longer to make decisions. According to McKinsey, if a clear inventory of data is not available, users can spend between 30 to 40 percent of their time searching, leaving less time to analyze or act upon it. **Discovering data** is key.

Step 2. Discover and Curate

2

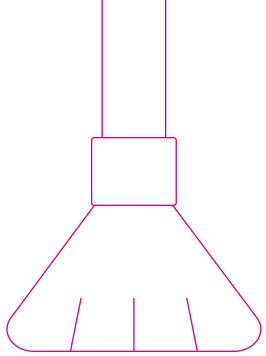
Curation comes next. Data curation involves managing data to make it both useable and useful for users and is essential to helping businesses users understand if a particular dataset will meet their needs. To do this, data must be tagged so users can easily find it, request and access it, and be able to comment on or discuss it. Although this may seem like a daunting task, intelligent discovery and auto tagging based on domain discovery can greatly accelerate curation and improve transparency about where data has originated, what has been done to it, and details about its entire lifecycle. Crowdsourced rankings and peer-to-peer collaborative comments, used in conjunction with metadata tags, help users know what exactly is in each dataset, and how—and how successfully—it has been used.

Unfortunately, curation is where most organizations are falling behind. According to IDC, only 25 percent of useful data created in 2018 was tagged. Even more significantly, only 13 percent of tagged data was used in analyses.

Checklist

- [ ] Do you know where all of your data lives?
- [ ] Do you have an automated, intelligent tool that can scan your entire enterprise and discover datasets?
- [ ] Can you produce a data lineage that automatically ties technical metadata to its business meaning?
- [ ] Do you have a mechanism for the user community to curate and comment upon data?

Step 3. Cleanse and Master



3

If the first two steps in this workbook are about acquiring knowledge regarding your data, Steps 3 and 4 are about ensuring that you can trust it.

After all, if you're urging business professionals within your organization to base decisions and actions on data, that data had better be trustworthy. Unfortunately, for the majority of businesses, it isn't: 60 percent of organizations say they are challenged by the quality and complexity of their data, according to IDC. Some insight into current data quality levels—delivered by running a quick, automated assessment—can help you pinpoint areas of focus aligned with your business priorities.

For data to be trustworthy, it has to be of high quality. It must be free of errors so that, once democratized, the results that your users get from analyzing it can likewise be trusted. As the IT saying goes, "Garbage in, garbage out." This is where **data cleansing** comes into play.

Data cleansing is the process of preparing data for analysis by removing or modifying data that is incorrect, incomplete, irrelevant, duplicated, or improperly formatted. It involves fixing spelling and syntax errors, standardizing data sets, and correcting mistakes such as empty fields and duplicate data points. Data cleansing is considered a foundational element of data democratization, as so-called dirty data will return untrustworthy results during analysis. According to McKinsey, if robust, automated, data cleansing tools and controls aren't in place, your users will spend 20 to 30 percent of their time doing manual data cleansing.

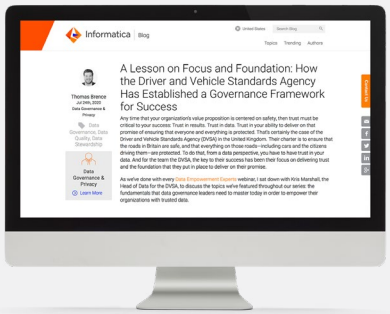
After data cleansing comes mastering, which is all about having a single source of truth. In your customer database, you don't want one record saying John Doe lives at 123 Main Street and has never purchased from you, while another record indicates that John J. Doe—with the same address and phone number—has made 50 purchases from you and is vocal about your company in online forums. This lack of consistency in datasets is exceedingly common. Organizations generally turn to master data management (MDM) tools to provide data consumers with a single trusted view into whatever dataset—customers, products, or operations—they want to analyze.

Step 3. Cleanse and Master

3

DVSA takes a "defensive strategy" to ensuring trustworthiness of data

The charter of the Driver and Vehicle Standards Agency (DVSA) in the United Kingdom is to ensure that the roads in Britain are safe, and that everything and everyone on those roads—including all vehicles and the citizens driving them—are protected. To do that, the agency needed to have trust in its data. But all its data was in silos, and was inconsistent in terms of driver names, license plates, vehicle numbers, and other critical data. "It's all very well to have data scientists lining up to use your data," says Kris Marshall, DVSA's Head of Data. "But we didn't have a single version of the truth." By taking a centralized approach to the reliability of its data—which meant giving power to the right people to make decisions—DVSA mastered its data with Informatica Master Data Management to be standardized, consistent, and trusted. DVSA used Informatica Axon Data Governance and Informatica Data Quality to ensure accuracy in the data, and to repair low-quality data before granting access to anyone.



Checklist

- [ ] Do you know if your data is of high or low quality?
- [ ] Can you easily visualize and report on this quality to executives and key stakeholders?
- [ ] Do you have a single source of trust for your critical data?

Step 4. Protect and Monitor

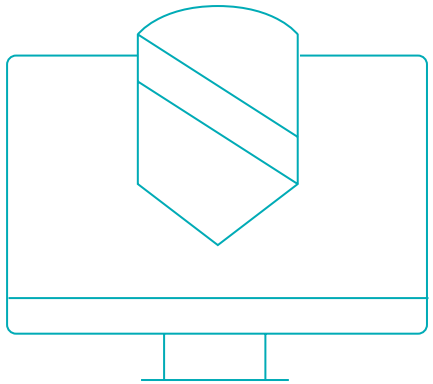
4

Whether we trust data sufficiently to use it ultimately comes down to two key concepts. We've covered the first one in Step 3, where we discussed how to measure and ensure data quality. In Step 4, we're concerned about privacy: do I (the user) legitimately have access and the proper rights to use this data responsibly?

Your users require clarity and transparency on both dimensions of trust.

Data privacy regulations such as the GDPR, HIPAA, the Federal Trade Commission Act, the CCPA and others have become much more commonplace to enforce personal data privacy compliance, with heavy fines and reputational damage for companies that violate them. However, according to IDC, only 57 percent of companies know where their sensitive data is stored. This presents a very real roadblock to your data democratization strategy.

So, ensuring that that sensitive data is reliably protected against even inadvertent misuse is an essential step before providing access to or democratizing the data. This step involves careful management of access rights to data and enforcing appropriate use policies, along with constant monitoring to make sure that protected data stays protected, while minimizing unnecessary risk exposure.



Step 4. Protect and Monitor

4

And here you also need to consider the source of sensitive data—the people that the data is about. Whether an individual consumer, a state resident renewing their driver's license, a participant in a medical study, or an employee's benefits information, people are increasingly sensitive about the appropriate use of their data: whether it's collected and handled responsibly, stored safely, and is used ethically, in line with stated purpose. Axios found that 88 percent of people are frustrated by the fact that they don't have control over how their personal data is used and wish the process of retrieving it from companies was easier and more transparent.

This is why data leaders should care, too, in order to maintain customer loyalty and preserve future revenue as a trusted data steward. To minimize frustration and respond with greater transparency to data subject access rights (DSAR) requests, it's best to take an intelligent, automated approach: one that ensures that only authorized users can access sensitive data for responsible use, and the organization can assess and reduce exposure when requested to do so to minimize unnecessary privacy risks.

Checklist

- [ ] Do you have an automated methodology to understand your data protection posture?
- [ ] Can you display privacy risk baselines to key stakeholders and executives for guided decisions?
- [ ] Can you handle DSAR requests efficiently to comply with privacy mandates at enterprise scale?
- [ ] Are you able to prioritize and remediate risk exposure by applying intelligent analytics?

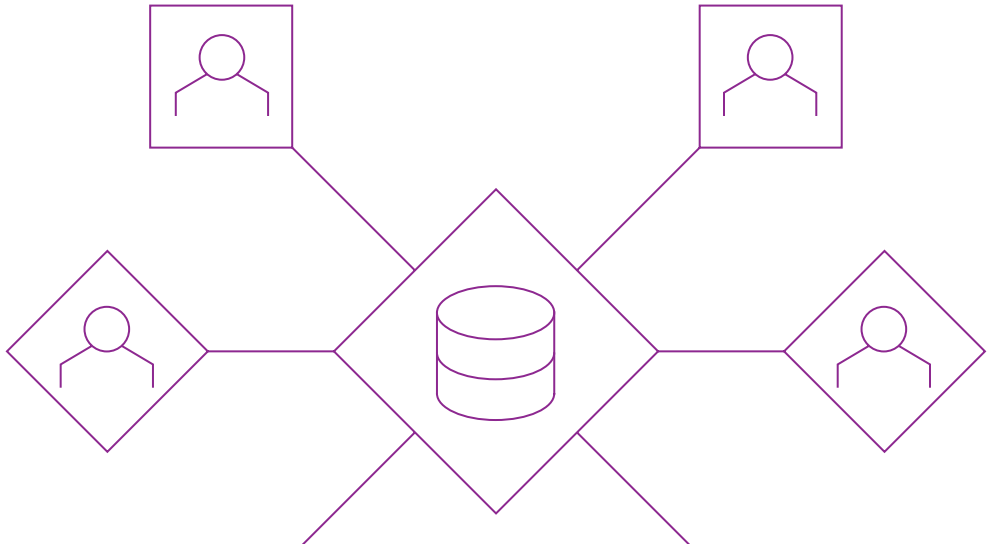
Step 5. Provision and Consume

5

We've talked about knowing your data (definition, discover) and trustworthiness (quality, privacy). Now we'll focus on useability. Steps 5 and 6 are about packaging up the data and getting it to the right place, the right people, in the right form so that it can be found and analyzed or used quickly and efficiently.

Useability is important because it's one thing to know the data exists, and that you can trust it. But that's just half the battle. If you can't get it to the right location in a format that allows it to be easily used, it's all for naught. According to IDC, data professionals spend 70 percent of their time searching for and preparing data, 20 percent of their time making sure it's being kept safe and protected and following governance policies, and only 10 percent of their time actually performing analytics. That's an enormous waste of time and resources.

Which means that organizations must have the ability to source, transform, blend, and then process data into high-quality, governed datasets before it can be shared and consumed by business users. This is called **data provisioning.**



Step 5. Provision and Consume

5

Data provisioning is the act of moving structured or unstructured data into a system, in a form that can easily be used in business intelligence (BI), analytics, and data visualization applications. Data provisioning frequently requires pulling data from different systems—both internal and external—and integrating it before loading it into a database, data warehouse, data lake, or other repository.

The sheer complexity of the data that resides virtually in modern data repositories today requires artificial intelligence (AI)-powered automation and allows modern data provisioning at enterprise scale. Tasks involved in the data provisioning pipeline, including inferring relationships about datasets, pattern recognition, and recommendation of alternative datasets, can be automated with user-friendly, easy-to-use tools that allow users themselves to transform data for self-service data integration and BI- or analytics-driven applications.

Checklist

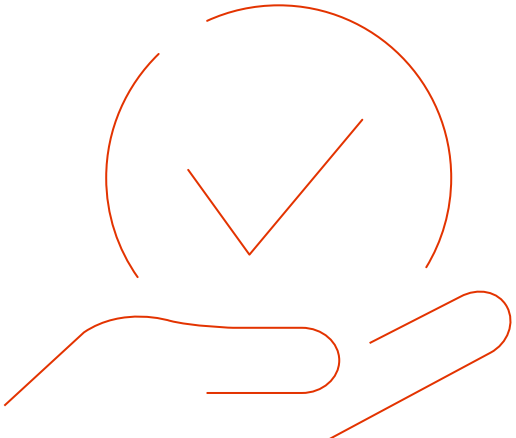
- [ ] Do you have tools to ensure data workers do not waste time provisioning data?
- [ ] Can data users easily request access to trusted data?
- [ ] Can you easily provision the requested information through an automated solution?

Step 6. Set Up Shop and Deliver Value

6

Now that you've created high-quality, trustworthy and protected datasets, you can publish and make them available for use.

What is a data marketplace? It's like an ecommerce store, but it's designed be used by your internal data consumers as they "shop" for data assets. A data marketplace provides an end-to-end experience for you to intelligently and automatically enable your data users to browse, shop, discover, and understand the context of data—and, if they're authorized, access it for responsible use that aligns to your governance policies.



Within a data marketplace, datasets are typically categorized based on various categories within your organization, such as finance, HR, or inventory management. Users can browse across those categories, or search using a specific keyword.

There are three stages to how a data marketplace operates:

Create and publish. These are the first four steps of the six-step process outlined in this workbook: the data has been discovered, it has been curated, cleansed, and mastered, and protected against unauthorized use.

Shop and checkout. This resembles popular online ecommerce websites, except users are browsing for datasets instead. And rather than look at customer reviews, they can reference crowdsourcing scores from other data consumers. They can even refer to quality scores and privacy policies associated with that dataset to give them context around the quality, trustworthiness, completeness, and sensitivity of the data. Once they've identified the datasets they want, they simply add them to their so-called shopping cart, and proceed to "checkout." Checkout triggers a workflow—either integrated with other, existing processes within your organization through service management tools, or through an automated tool—where the owner of the dataset is asked to authorize use of the dataset for that particular user.

Step 6. Set Up Shop and Deliver Value

6

Fulfill and track. Once authorized, that's when metadata intelligence gets leveraged. Because the dataset defines what the final data structure will look like, you can automatically generate integration mappings and provision the necessary integration jobs to package up the data and send it to the location of the user's choice.

The more automation you can apply to the provisioning needed to fulfill marketplace data, the more broadly your marketplace will be able to address consumers' data needs on a self-service basis. If ever there was a time for AI and automation, it's now.

The payoff: if you can get trusted data into the hands of your people, they can do amazing things. Companies that empower employees to consistently use data as a basis for their decision-making are nearly twice as likely as others to report reaching their data and analytics objectives, according to McKinsey in "How Leaders in Data and Analytics Have Pulled Ahead."

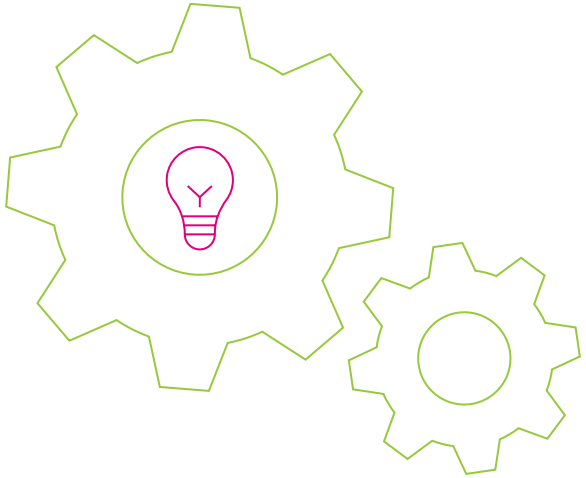
Checklist

- [ ] Does your team have the ability to create and publish?
- [ ] Can your data consumers shop and checkout?
- [ ] Can you determine whether users have the right to access data they request?
- [ ] Can your team fulfill requests and track usage automatically?

Conclusion

Data Democratization Requires Intelligence and Automation

Data Democratization Requires Intelligence and Automation



To successfully tackle and complete each of the six strategic steps for data democratization, you need a solution that's powered by intelligence and automation. As IDC's Stewart Bond points out, "organizations that apply intelligent automation to enable their people and processes with data, have a better opportunity to succeed within the scale of modern data environments."

The risks of not doing it right are very real. Even Fortune 500 companies face challenges when making data widely accessible to their people. All too often, organizations lack the protocols needed to accurately track who is using data and for what purpose. For example, a large defense contractor found that its more than 600 data analysts and scientists were spending 14% of their day searching for information to complete simple tasks such as creating reports, calculating metrics, and testing models.

This is changing rapidly. Eight in 10 organizations intend to move from an IT-centric to self-service data model by 2024. Although this requires both serious money and work, it comes with a significant payoff. McKinsey found that within three years, organizations investing in automating data self-service consumption like a data marketplace can expect to save anywhere from 30 to 50 percent of their data-management costs.

By building a new self-service model that supports data democratization, you can empower teams across your organization—from the executive suite to marketing to manufacturing—to drive productivity, efficiency, and the effective use of data.

But this worthy goal of governed trusted data democratization can be achieved only with strong, scalable data governance and privacy. And the only way to get there is through AI- and machine learning-powered intelligent automated data governance and privacy.

Learn More

Informatica Democratize Data Access Across the Enterprise

For more information, download the executive brief.

READ MORE

About Informatica

Digital transformation changes expectations: better service, faster delivery, with less cost. Businesses must transform to stay relevant and data holds the answers.

As the world's leader in enterprise cloud data management, we're prepared to help you intelligently lead—in any sector, category or niche. Informatica provides you with the foresight to become more agile, realize new growth opportunities or create new inventions. With 100% focus on everything data, we offer the versatility needed to succeed.

We invite you to explore all that Informatica has to offer—and unleash the power of data to drive your next intelligent disruption.

Worldwide Headquarters
2100 Seaport Blvd, Redwood City, CA 94063, USA
Phone: 650.385.5000
Fax: 650.385.5500
Toll-free in the US: 1.800.653.3871

informatica.com
linkedin.com/company/informatica
twitter.com/Informatica

CONTACT US

IN19-1020-3991

Informatica®