Your Organization's Guide to Data Maturity

Table of Contents

datacamp

1

The state of data-driven digital transformation

In 2012, Harvard Business Review dubbed data science the "sexiest job of the 21st century". In it, D.J. Patil and Thomas H. Davenport outlined the nascent set of skills and attributes that make up a data scientist, and how leveraging data and analytics will be the main competitive advantage that sets organizations apart from their competitors.

Now, almost a decade later, achieving digital transformation, becoming data-driven, and scaling data science capabilities is on every organization's mind. According to a 2021 NewVantage Partners survey, 99% of firms are reporting active investment in data science and machine learning. However, fewer than 30% of organizations are experiencing transformational business outcomes as a result of these investments, and only 24% of them claim they have created a data-driven organization. Similarly, research from BCG shows that more than 80% of companies are accelerating their digital transformation plans, but only 30% of them are achieving their stated objectives.

The path to becoming a data-driven organization is a long and arduous process (Harvard Business Review), as it requires investments in scaling various levers such as infrastructure, people, tools, and more. Organizations also sit on a spectrum of data maturity and depending on where they sit on this spectrum, need to allocate their focus to some levers over others.

The north star for any aspiring data-driven organization is to become data literate, where everyone has the access, and skills they need to work with data and make data-informed decisions. This ultimately means the democratization of data and skills across a range of roles and the maturation of a data culture.

In this guide, we outline a framework for evaluating, and scaling data maturity throughout the organization, define the various data maturity stages an organization goes through, and draw a path of initiatives organizations can take to achieve data literacy.

2

The five data maturity levers

When scaling data science throughout an organization, there are five key levers one should consider to go from one data maturity stage to another. These levers are the key elements of our **IPTOP** framework for scaling data science, and they can be defined as such:

- **Infrastructure** — entails creating a data infrastructure that ensures data is collected, discoverable, reliable, understood, compliant, and actionable throughout the organization.
- **People** — entails both forging a data culture where everyone understands the value of data and the importance of continuous learning, as well ensuring that everyone has the skills to work effectively with data.
- **Tools** — which refers to the tools data practitioners across all skill levels use, as well as the development of frameworks that enable easier work with data throughout the organization.
- **Organization** — which refers to how data talent is organized, and adopting organizational models that promote scalable data science throughout the organization.
- **Processes** — the workflows that data experts and teams adopt to make their work more predictable and collaborative, and to ensure alignment with business objectives.

Infrastructure and people are the base foundation supporting the other levers. Without data culture, skills, and access to data, organizations won't be able to leverage data tools, scale data processes, and organize data talent.

3

Data maturity is a spectrum

The path to data literacy is comprised of four data maturity stages. It's important to think of data maturity as a spectrum, and that some organizations may find themselves relatively more mature in one element of the IPTOP framework and less mature in another. With that in mind, we can break down the stages of data maturity through the number of people in an organization that rely on data to do their best work, and who have the skills and access to make data-driven decisions.

It's important to note that this does not mean that everyone needs to code, as there are various data personas found in every data-driven organization who possess a completely different relationship with data to do their best work. For example, data leaders and consumers are often non-technical roles that need to be data literate in order to be conversational with data experts and identify where data can be used to answer a business question. Data engineers, on the other hand, are highly technical individuals who are responsible for scaling and maintaining data infrastructure throughout the organization.

For a deep-dive into all eight data personas found in every data-driven organization, download our white paper on the [L&D Guide to Data Fluency](#).

Data Reactive

No one accesses or uses data in their daily work. An organization rarely reports or presents data.

Data Scaling

Very few people in an organization have the skills and access they need to analyze, report, and confidently present data insights.

Data Progressive

Every team has at least one data practitioner who can analyze, report, and present their data regardless of their role.

Data Literate

Everyone knows how to access the data they need to do their job and make data-driven decisions.

4

What data reactive looks like

In data reactive organizations, no one accesses or uses data in their daily work, and a company rarely reports on or presents data. There is no data culture or strategy to support maturing data competencies.

Infrastructure

Data is collected on an ad hoc basis with disparate tools, and there is no trusted, centralized data storage throughout the organization.

Tools

There are a variety of ad hoc legacy tools for working with data that are rarely leveraged.

Processes

Any form of data work is done on an ad hoc basis and there are no established processes for working with data.

People

No one has the skills to work with data on a daily basis — and the organization does not have nor does it prioritize developing a data culture.

Organization

There is no data team and no data experts throughout the organization. A lack of data strategy hinders the allocation of resources for data talent.

At the data reactive stage, organizations suffer from low maturity in infrastructure and people, ultimately limiting their maturity ceiling in terms of tools, organization, and processes. On the infrastructure side, a data reactive organization collects data on an ad hoc basis and stores it in siloes. The data often sits in departmentally owned flat files, spreadsheets, or legacy tools. There are no data quality standards to support data linkage between departments, and there is minimal data strategy that incorporates scaling data infrastructure.

On the people side, data reactive organizations are characterized by cultures that value experience and gut instinct over data-driven decision making, there is no dedicated data talent working on scaling data maturity, the majority of the organization does not value the opportunities of employing data analytics in their daily work, and learning and development for data skills is not a priority on the organizational roadmap.

"Building a data culture is not just an option, it is business-critical. Literally 100% of data and analytics investments depend on having a Data Culture."

Sudaman Thoppan Mohanchandralal, Chief Data Officer at Allianz Benelux

The importance of scaling people

Arguably the most important lever to scale when scaling data maturity in the early stage of data maturity is people, specifically data culture and skills. According to the same NewVantage Partners survey covered earlier, 92% of organizations believe that the biggest impediment towards becoming data-driven is people and culture. This is because regardless of technological investments in data infrastructure and tools, people drive the success of data and analytics initiatives (Deloitte).

Moreover, scaling data skills and culture means that there is organizational alignment and excitement around the importance of data initiatives (McKinsey). This excitement will be the lifeblood that sustains future initiatives when scaling data maturity and will be the catalyst of growth for organizations on their path to data literacy.

The path to data scaling

The path to scaling starts by focusing exclusively on the two base levers of the IPTOP framework: **infrastructure** and **people**. The initiatives laid out below are centered around forging data culture and talent, the development of a data strategy that takes into account scalable data infrastructure alongside learning and development, and ultimately setting the stage for greater growth in the long term.

People

- **Prove the value with a proof of concept** — In order to create excitement around data initiatives and to drive executive sponsorship of organization-wide data culture change, organizations should look to develop a proof of concept that showcases the potential return on investment of data projects. The goal here is to showcase a quick win that is simple and robust in order to show the potential of wider investments in data. These analytics low-hanging fruit can range from developing a simple customer churn model that showcases the potential of optimized marketing spend, developing a rudimentary dashboard that provides descriptive analytics for finance departments, or even leveraging A/B testing on email subject lines when managing email campaigns to optimize open rates.

- **Build strong executive support** — According to McKinsey, one of the key differentiators between organizations who succeed at scaling data initiatives versus those who lag behind is strong executive support and ensuring senior-management participation in analytics activities. More importantly, organizations with high-performing data initiatives report higher direct sponsorship of data initiatives by the CEO. Developing a top-down approach for driving excitement around data initiatives is essential to sustaining data culture change. Practically speaking, this could begin with public championing of successful proof of concepts, developing and articulating an organization-wide data strategy, ensuring senior management are putting data on their departmental agenda, and hiring data talent to champion and enact change (Harvard Business Review).

- **Develop a data strategy that puts learning at the center** — Once executive support is acquired and an organization is aligned on the importance of scaling data science capabilities, leaders should start developing a data strategy that puts learning and development at the center. A holistic approach that integrates business strategy, data infrastructure, governance, and implementation outcomes is essential to drive success (Forbes). However, a key component of scalable data strategies is learning and development as it will empower executives to navigate complex data landscapes with improved data literacy, lay the groundwork for sustainable organization-wide data maturity in the long run with data skill assessments, evolve the skill set of subject matter experts to meet future data challenges, and be a key component of any data culture program (Deloitte).

- **Hire the appropriate data talent** — As excitement begins to build around data initiatives, organizations should start looking to hire data experts to start working on key pilot data projects that could have transformational business outcomes. organizations must hire data talent that focuses on the foundations that enable scalable data science in the long run. A key anchor for organizations to leverage is the Data Science Hierarchy of Needs developed by Monica Rogati, which articulates that the path towards using machine learning and advanced data science at scale starts with fostering a solid foundation in terms of data collection, storage, and transformation. While no organization is the same, hiring data engineering talent should be top of mind for organizations looking to launch their path into data literacy (Quantum Black).

Infrastructure

- **Develop a data strategy with scalable infrastructure in mind** — In order to enable data access at scale for the long run, organizations need to also center their data strategy around scalable data infrastructure. Depending on the size and complexity of the organization, such a strategy could mean defining how and which data will be collected, outlining initiatives for modernizing IT infrastructure (McKinsey), scoping data architecture and its requirements (Deloitte), thinking about data quality, identifying vendors for cloud migration, and more. In the next data maturity stage, we will outline concrete steps organizations can take to scale their data infrastructure.

5

What data scaling looks like

In data scaling organizations, very few people in an organization have the skills and access they need to analyze, report, and confidently present data insights. There is minimal data culture with very few champions and very few skilled practitioners, and there is no organization-wide access or trust in data.

Infrastructure

Only a few key experts understand how data is accessed in the organization. There is no organization-wide access to or trust in data.

Tools

There are mostly legacy tools for working with data, with ad hoc use of modern tooling for data work.

Processes

Very few, ad-hoc and limited data processes exist in siloed teams.

People

There is minimal data culture with very few champions believing in the importance of leveraging data, and few data experts that are working on data initiatives.

Organization

While a data strategy is in place, the organizational structure of data talent is still undefined and change management is still taking place.

At the data scaling stage, organizations have made the first steps by creating a data strategy that aligns data science with business objectives, is centered around data culture and skills, and incorporates scalable data infrastructure. Moreover, they have some data talent in place to drive some strategic projects. Having made these first steps, data scaling organizations will need to start strengthening maturity of tools, organization, and processes.

Infrastructure for data scaling organizations is not yet mature, as only a few key experts understand how data is accessed, and there is still no widespread trust in data. This means that there is no centralized data storage that allows easy access for data experts and enthusiasts alike, datasets are still siloed, and data experts face data quality issues that lengthen the data cleaning process and stand in the way of joining datasets across various departments. More importantly, some departments have gotten a head start over others in terms of accessing and utilizing data to push key strategic pilot projects.

On the people side, data scaling organizations can be characterized as having minimal data culture with very few people who believe in the importance of data, evidenced by not having the skills or mindset to make data-driven decisions. This can be manifested as a data culture that is still undefined, and whose exact components are not yet articulated, and where self-motivated learning is king. Moreover, as data initiatives start gaining traction alongside data experts, change agents and data champions will start to crop up in forward-thinking departments (DataIQ).

As for the remaining components of data maturity (tools, organization, and processes), organizations at this stage of maturity do not have a modern data tooling stack. Instead, most data work is done in legacy tools such as spreadsheets, and only a few key data experts leverage advanced data tooling. More importantly, organizations have still not developed an organizational model for data talent, and senior leadership is not yet spearheading or sponsoring change management. Finally, only a few ad hoc data processes exist on siloed teams, with no organization-wide norms set in place to work with data or collaborate with data experts.

Strengthening foundations will be key

In this stage of data maturity, organizations will still need to focus on the foundations. As discussed earlier, people are arguably one of the most important levers when scaling data maturity, in terms of their skills and culture. Scaling people will not require just organizational alignment around the importance of data science or data initiatives, but a behavior and mindset change that trickles down from the top to the bottom of an organization (Harvard Business Review).

> "Data Science is a way of doing things, and not a thing to do [...] Data Competency is a way of thinking, and not a thing to think about"
>
> **Kirk Borne, Chief Data Scientist at Booz Allen Hamilton**

As such, organizations need to articulate and define a data culture by defining important metrics for business initiatives, defining and incentivizing new behaviors (e.g., reporting on project updates with data points), developing a wider learning and development strategy to shore up data skills, and rewarding data champions and change agents. Strengthening infrastructure is the second foundation for data to permeate throughout the organization, which will allow organizations to scale data access. This includes centralizing data using a centralized data storage solution, defining data access structure for high-impact teams, and establishing data governance and quality mechanisms and policies.

The path to data progressive

Infrastructure

- **Start creating a single source of truth by centralizing data storage** — A key component in scaling access to data that is discoverable, reliable, understood, compliant, and actionable throughout the organization, is centralizing data storage and creating a single source of truth for the organization (McKinsey). While just centralizing data storage won't solve an organization's data quality woes, it's an excellent first step in avoiding siloed datasets that do not adhere to organization-wide data quality standards. Centralizing data storage is most commonly done by leveraging centralized data storage solutions like data lakes and warehouses from cloud providers such as Amazon Web Services, Microsoft Azure, Google Cloud, Snowflake, and more.

- **Start scaling data quality** — One of the main obstacles towards organization-wide trust in data is data quality (Monte Carlo). Moreover, data quality issues hamper productivity as data experts spend excessive time on data quality issues rather than advancing an organization's data maturity (McKinsey). As such, organizations looking to mature their infrastructure should make data quality and governance a key aspect of scaling infrastructure on the path to data progressive. This includes developing data quality standards for various departments, creating documentation around accessing data, investing in data quality monitoring mechanisms, and iteratively increasing the scope of data governance initiatives.

People

- **Clearly define what a data culture is and what it isn't** — As leadership starts to communicate the importance of data culture, they must also define what a data culture entails. In a data-driven culture, organizations set clear expectations on how to use data responsibly, executives exemplify data-driven behavior by arguing and persuading with data, and a data-driven mindset becomes a habit (Tableau). As such, organizations should start defining and tracking important metrics around projects, encourage agility and the use of data, tie data-driven decision-making to business outcomes, and drive continuous education.

- **Create an upskilling strategy that's made for the long-term** — A key component of creating and reinforcing a data culture is making sure everyone has the skills they need to work with data effectively to do their role. At this stage, organizations should set the stage for an organization-wide data upskilling strategy that is inclusive and takes into account all types of roles. This begins with identifying relevant data personas in the organization, assessing current skill sets and identifying skill gaps, and most importantly, aligning the upskilling objectives with business strategy (DataCamp).

Learn how AXA XL is using DataCamp to upskill its people

Learn more >

- **Reward the change agents** — A data scaling organization is typically characterized by the presence of data champions and change agents who are playing an active role in evolving the status quo. Organizations should reward these change agents by increasing the visibility of their work, promoting and empowering them to be more cross-functional, and evangelizing the successes they've led across the organization. This further catalyzes culture change and increases the adoption of data-driven buy-in throughout the organization (Tableau).

Tools

- **Modernize your data stack** — As data becomes centralized and providing scalable access becomes top of mind, giving access to data tooling that is inclusive to relevant data personas for your organization will be paramount. This could include open-source programming languages like Python and R, business intelligence tools such as Tableau or Microsoft Power BI, easy access to a SQL database interface, and more.

Organization

- **Define how data talent is organized** — With increased hiring of data talent (i.e., data scientists, data engineers, data analysts, etc.), a key component of scaling data maturity is defining an organizational model for data experts. Depending on the size and the objectives of an organization, data talent can be organized under a central data team or can be distributed across business units. The pros and cons between the centralized data team model and the decentralized model vary. The centralized model allows the data team to function as a center of excellence, with the ability to push the data agenda throughout the organization. Moreover, it allows collaboration and knowledge sharing between data experts who are working on completely different problems. However, a drawback is that it could limit collaboration and alignment with subject matter experts, and risks isolating the data team as a support function. The decentralized model, on the other hand, allows data experts to be directly embedded in business units. This allows better cross-functional alignment between business stakeholders and data scientists. However, this also means that it's more difficult to move data experts between teams, and data experts are managed by business leaders.

Processes

- **Create processes for collaboration between data talent and business units** — Depending on the model chosen for organizing data talent, organizations should start thinking about creating scalable processes to enable better collaboration between data scientists and subject matter experts and to increase the productivity of data experts. This could start with a ticketing system, creating a Quality Assurance process before publishing data products, adopting agile project management methodologies, or creating template project structures.

Centralized Model

Benefits

- ✓ Data science functions as a center of excellence
- ✓ Promotes collaboration and knowledge sharing
- ✓ Data Science Manager has domain knowledge
- ✓ Incentivizes consistent technology stack and better tooling

Drawbacks

- ✗ Limits coordination between data science and other stakeholders
- ✗ Risk of misalignment between data science and Business Units
- ✗ Risks isolating data science as a Support Function

Decentralized Model

Benefits

- ✓ Each team has a dedicated data scientist
- ✓ Cross-functional alignment
- ✓ Data science has a more natural "seat at the table"
- ✓ Fewer dependencies across teams

Drawbacks

- ✗ Harder to move data science resources between teams
- ✗ Manager of the team may not have domain knowledge
- ✗ Harder for data scientists to collaborate and drive long-term projects

6

What data progressive looks like

In data progressive organizations, every team has at least one data literate employee who can analyze, report, and reason with data regardless of role. Data is accessible but not easily discoverable. Data is seen as a strategic asset, but there is no organization-wide data literacy.

Infrastructure

Data is accessible, and data infrastructure is maturing. However, data is not readily discoverable, compliant, understood, or actionable throughout the organization.

Tools

Modern tooling is accessible, but there are no frameworks that democratize working with data.

Processes

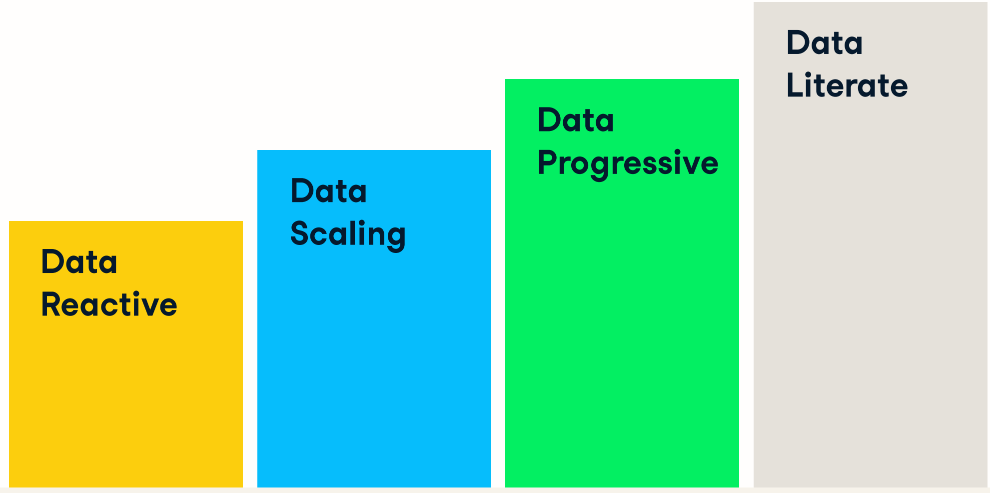
Mature processes for high data competency teams, but are yet to be democratized for the entire organization.

People

Data is seen as a strategic asset and there is alignment around its importance. However, organization-wide data literacy is lacking and data training is limited to pockets of the organization.

Organization

The data team is set in place, but their impact is siloed and data talent is seen as a support function.



At the data progressive stage, organizations have made strides to elevate data maturity. This translates into an environment where there is organization-wide buy-in over the importance and strategic nature of leveraging data, and where data is largely accessible. There is an established organizational model for data talent, and modern tooling is accessible across the board for various data personas. However, the final set of challenges here are all about democratization—empowering everyone to make data-driven decisions, freeing up data talent from becoming a support function, and focusing instead on highly strategic projects that capture value and raise the organization's data maturity (O'Reilly).

On the infrastructure side, data is accessible and data infrastructure is maturing. However, there is no easy, scalable way to easily discover who owns a particular dataset and what changes have been done to it, nor to find a clear explanation of dataset metadata and other key components of data discovery. Moreover, scalable organization-wide data quality standards are still lacking, which hurts overall trust in data. Finally, operationalizing machine learning and data science is still nascent, and many tools aimed at productionizing, monitoring, and refining models in production have yet to be adopted.

On the people side, strides in data culture have led to organization-wide buy-in on the strategic importance of data. However, data skills training is still departmental. Insufficient organization-wide data literacy slows down innovation and limits productivity (Accenture). Most importantly, it hinders the adoption of data science as a methodology for solving problems, rather than a skill to acquire. This ultimately results in an organization where there is no common data language, an over-reliance on data talent to provide data insights even if it's basic reports, and where people do not feel empowered to make data informed decisions.

As for *tools, organization, and processes*, data progressive organizations have a modern tooling stack but have yet to create scalable frameworks that further democratize working with data. This means that working with relatively advanced tools has a high barrier to entry for non-data experts. The data team has an organizational model (centralized or decentralized), but data talent is limited to being a support function for other business units. This limits the data team from advancing the data agenda by pursuing high-impact projects that scale throughout the entire organization. Data progressive organizations also have high maturity processes for streamlining data talent workflows, but ways of working around operationalizing machine learning and data science are still nascent, and processes for scalable data work are not inclusive to the remainder of the organization.

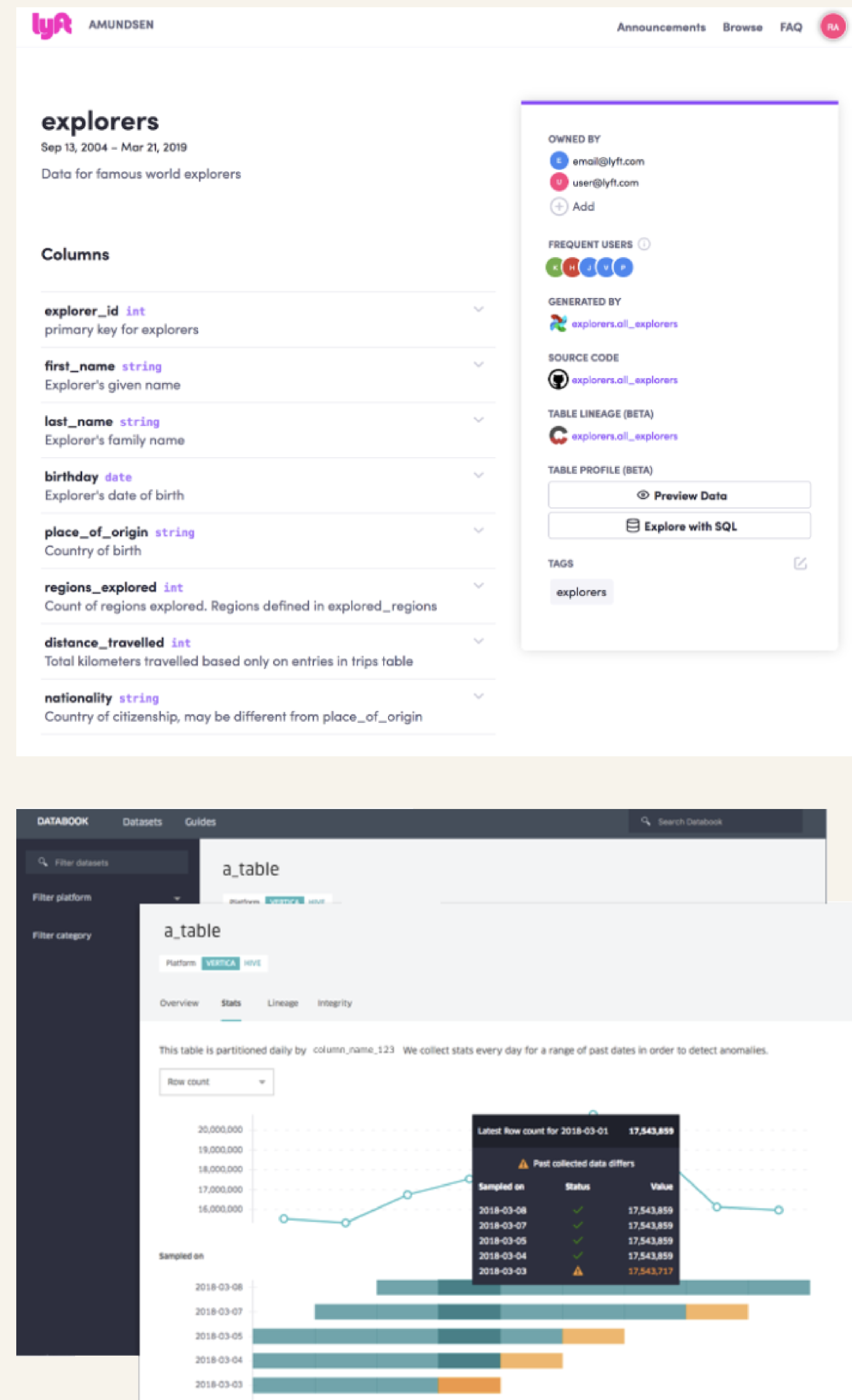
The path to data literacy requires a multipronged approach

The arrival at data literacy is arguably the most difficult step to make for organizations. At this point, organizations are likely to have scaled their infrastructure to process terabytes of data daily, hired, organized, and derived value from data talent, and enabled access to modern tooling.

The difficulty is scaling data science to become an organization-wide methodology for solving business challenges, and where data and data skills are democratized. To reach this point, organizations should adopt a multipronged approach that tackles all the elements of the IPTOP framework.

Infrastructure

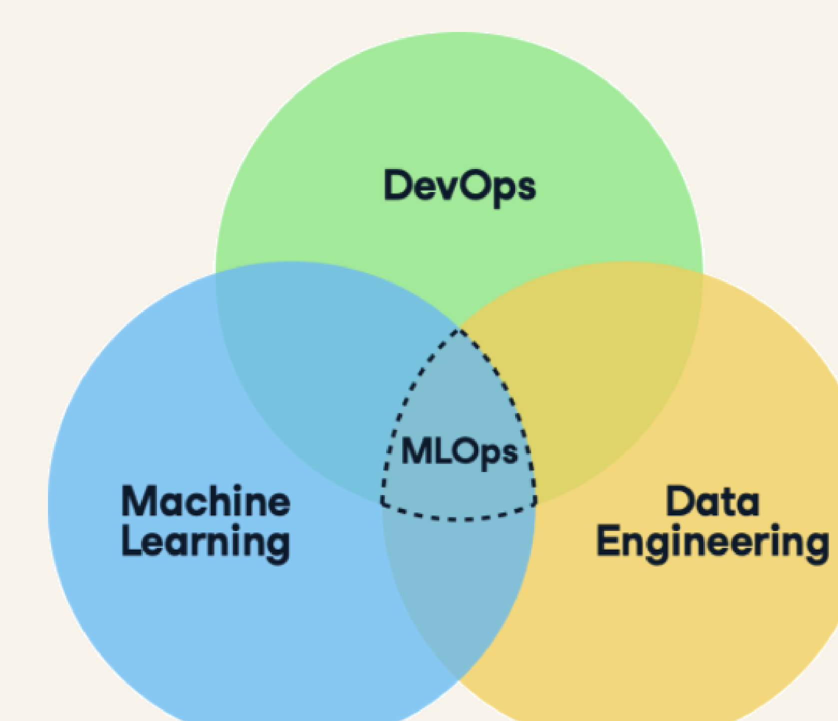
- **Start investing in data discoverability** — As data infrastructure scales and the architecture of [modern data infrastructure] becomes standardized, a key component of democratization and scaling data science will be ensuring that data is not only collected and accessible, but trusted, reliable, actionable, and easily discoverable. This is where data discovery tools come in. Data discovery tools were born out of the growing pains of data literate organizations. For example, Lyft developed its open source discovery tool [Amundsen], based on the challenges associated with scaling the amount of data being collected. Airbnb developed [Dataportal] in order to make sense, and tackle the challenges with an increasingly complex data landscape. Data discovery tools provide powerful search functionality, a deep overview of dataset ownership, lineage, metadata, and programmatic functionality to scale data quality checks and descriptions ([Edmunds]).



Lyft's Amundsen, and Uber's Databook in action

- **Strengthen data quality and governance** — In order to create trust and alignment around data, organizations looking to democratize data should look at data governance as a key pillar for facilitating that. As such, organizations should invest in scaling their [data quality strategy and processes]. This means investing in data governance platforms, treating data as a product to which it abides by software engineering best practices like data service level agreements ([Monte Carlo]) and creating data quality certifications for datasets ([Airbnb]).

- **Invest in operationalization capabilities and tooling** — As machine learning models become more commonplace —whether in business processes or in customer-facing products—operationalizing these models will require continuous monitoring, integration, versioning, and explainability. MLOps is still emerging as a field and goes beyond just tools and frameworks to adopt. The tooling space for operationalizing models in production is gaining steam and will allow organizations to scale deployment of data products and machine learning models in production ([INNOQ]).



People

- **Roll out organization-wide data training fit for all data personas —** We've demonstrated that people are the cornerstone of data maturity across stages. Organizations looking to drive data literacy at this stage face the challenge of making data a default consideration in most decisions and actions taken. One of the main blockers organizations face is insufficient data literacy (Deloitte), which limits the ability of non-experts to become independent when seeking and creating data insights. When rolling out organization-wide training, the best practice is to create personalized learning paths for different data roles. For example, Airbnb created Data University, a program aimed at upskilling the entire organization on data skills, with content curated for specific data teams in their Data U Intensive program. Cohorts were able to achieve a 30% sustained increase in SQL usage. Bloomberg uses DataCamp as part of a blended learning program, which has sustained a 561% increase in activities on their proprietary Jupyter Notebook environment (DataCamp). In a data literate organization, learning paths run the gamut from general data literacy, to data science, machine learning, and data engineering, and more. This ultimately enables the adoption of a common data language and reinforces data science as a habit and methodology for approaching business problems.

- **Assess, measure, and track data skill competencies over time —** A key component of scaling data literacy is tracking skill development and using assessments to steer learning paths (Deloitte). Assessments provide individuals with insights into their data skills maturity, and will guide personalized learning paths for them based on their strengths, weaknesses, and necessary data skills for their particular role. At an aggregate level, assessments will provide organizations an overview of their data maturity skill matrix, which acts as a balance sheet of the company's data skills.

DataCamp Signal™ helps organizations to pinpoint their strengths and skill gaps at scale.

- **Reward a culture of continuous learning —** Data science is not merely a skill to learn, but a methodology for solving business problems. As a field, data science is continuously evolving, with new techniques, frameworks, tools, and best practices being developed (Datafloq). Organizations should reward a culture of continuous learning if they want to build and sustain data skills across data roles (McKinsey). This not only includes creating a safe space for active learning and development, but incentivizing novel, data-driven approaches for solving problems and promoting learning by doing.

Tools

- **Develop frameworks to democratize working with data —** Providing access to tooling is only one part of the equation in democratizing data. The next frontier is lowering the barrier to working with data with frameworks. For example, the data team at Airbnb developed an R package conveniently named Rbnb, which provides a streamlined way to move data between different places in their data infrastructure, branded visualization themes, R Markdown templates for different types of reports, and custom functions to optimize different parts of data workflows. At DataCamp, we have R and Python packages that abstract querying data from our data warehouse with a few lines of code. Frameworks enable everyone to adopt more advanced data tooling, and standardize the data workflow for data experts throughout the organization.

Organization

- **Consider moving to a hybrid organizational model** — As data competencies mature and more data experts are developed within the organization, the next step is to organize data talent under a hybrid model. The hybrid model combines the centralized and decentralized data science team. In a hybrid model, a central data team functions as a center of excellence that drives data democratization efforts like creating frameworks and standardizing the data tooling stack throughout the organization, all while embedding data scientists into business functions to create better alignment between data science and business units. However, it's important to note there is a spectrum of hybrid models sitting between the two ends, and that organizations should find a sweet spot that suits their size, objectives, and data strategy (Altexsoft).

Processes

- **Create and reward processes that put collaboration and democratization at the center** — A key mark of data literate organizations is sharing tools, knowledge, and insights. These organizations have created processes that simplify access to data insights, making it easy to collaborate with data experts and work with data. For example, LinkedIn developed DataHub, a data discovery tool that also allows the organization to share data insights, charts, and dashboards. Airbnb developed Knowledge Repo, a platform that makes it easy to publish reports and to surface data insights. Netflix created processes and templates for accessing and working with Notebooks throughout depending on the type of problem being tackled. These processes enable easier access to data knowledge, and are centered around simplifying work with data.

The Hybrid Model

Benefits

- ✓ Data science can function as a center of excellence.
- ✓ Data science can drive common tech stack, tooling, frameworks, and standardization
- ✓ Data science can collaborate and align on organizational goals.
- ✓ Better alignment between Data science and business units

Drawbacks

- ✗ Risk of mismatch of expectation leadership of Data science and business unit.
- ✗ Everyone has at least two teams.

A blueprint for data literacy

	Infrastructure	People	Tools	Organization	Processes
Data Reactive	Data is collected on ad-hoc basis with disparate tools, and there is no trusted, centralized data storage ✓ **Develop a data infrastructure strategy**	No one has the skills to work with data — the organization does not have a data culture ✓ **Prove value with proof of concept** ✓ **Build executive support** ✓ **Center learning around data strategy** ✓ **Invest in data infrastructure talent**	Variety of ad-how legacy tools to work with data that are rarely leveraged ✓ **Prioritize infrastructure, and people**	There is no data team, and no data strategy to support it ✓ **Prioritize infrastructure, and people**	Any data work is done on ad-hoc basis and there are no processes for working with data ✓ **Prioritize infrastructure, and people**
Data Scaling	Only a few key experts understand how data is accessed in the organization. There is no organization-wide access to, or trust in data ✓ **Centralize data storage** ✓ **Establish data governance and quality policy** ✓ **Define data access for high-impact teams**	Minimal data culture with very few people believing in the importance of data or having the skills to work with data ✓ **Reward change agents and champions** ✓ **Define and outline data culture** ✓ **Set the stage for organization-wide upskilling**	Mostly legacy tools with ad-hoc use of modern tooling ✓ **Provide access to inclusive modern tooling** ✓ **Align tooling with infrastructure strategy**	Data strategy in place — with no centralized data team or embedded expertise set in motion ✓ **Define data team organizational model**	Very few, limited data processes exist in siloed teams ✓ **Define data team processes with other business units**
Data Progressive	Data is accessible, and data infrastructure is maturing. However, data is not easily discoverable, compliant, understood, or actionable. ✓ **Democratize data access with data discoverability and management tools** ✓ **Strengthen data quality and operationalize data**	Data is strategic, but underutilized throughout the organization. Organization-wide data literacy is lacking, and data upskilling is still limited. ✓ **Roll out organization-wide data upskilling fit for all data personas** ✓ **Assess, track, & reward skill development** ✓ **Start innovating with data**	Modern tooling accessible, however limited data democratization hinders value. ✓ **Develop frameworks to democratize data and lower barrier to entry to working with tools**	Data team set in place, however impact is limited to requests and analysis ✓ **Develop a hybrid model of embedded and centralized, to drive data strategy and expand value**	Mature data processes for high data competency teams only ✓ **Develop scalable data processes through organization by centralizing shared insights, promoting collaboration, and lowering barrier to entry**
Data Literate	Data is collected, discoverable, reliable, understood, compliant, and actionable throughout the organization ✓ **Innovate and automate infrastructure processes** ✓ **Monitor data products in production**	Everyone has the skills necessary to work with, and understand data. Continuous learning is part of the data culture. ✓ **Keep learning central to organization success** ✓ **Measure skill matrix development through organization**	Modern tooling and frameworks enable higher adoption and easier data driven decision-making ✓ **Refine frameworks and contribute to open-source community** ✓ **Invest in collaborative tooling**	Organizational model for scalable data science ✓ **Refine ad-hoc organizational models and enable further democratization**	Data processes to scale collaboration and efficiency ✓ **Center collaboration at the heart of data processes**

Data literacy is driven by people

Ultimately, a data literate organization is where everyone knows how to access the data they need to do their job. This could mean marketing managers accessing business intelligence dashboards to make decisions around campaign spend, executives surfacing data insights on a knowledge repository to make data-driven decisions, business analysts leveraging SQL to streamline complex excel workflows, and data experts leveraging the latest and most advanced data tools and techniques to further democratize data and push innovation throughout the organization.

What is common between these individuals is a mindset and buy-in around the importance of leveraging data for making decisions, and for solving problems. This is why building data skills and culture is paramount for any data maturity initiative to scale regardless of where the organization sits on the data maturity spectrum.

How DataCamp supports organizations on their path to data literacy

DataCamp's proven learning methodology provides a cyclical process for learning and retention. This learning methodology enables learners across the data literacy spectrum to assess their skills and identify gaps, develop a learning plan based on these gaps, practice skills, and apply them in a real-world setting. Experienced data scientists can upskill on new techniques in their target domain, and domain experts can learn the fundamentals of data literacy and data science.

Assess

Effective learning starts with understanding skill gaps and strengths. With [DataCamp Signal™](), learners can understand specific skill gaps they have across various topics and tools. From data literacy assessments like understanding and interpreting data to programming and machine learning assessments in R, Python, and SQL, our 10-minute adaptive evaluations provide learners with personalized skill gaps and learning paths to address their skill gaps.

Learn

DataCamp's growing course library houses more than 350 expert-led, hands-on courses across various technologies and domains for all data skills and levels. Learners can hit the ground running with our learn-by-doing approach—our bite-sized videos and interactive coding exercises allow them to start working with their preferred tool and topic right in the browser.

Practice

The next step in DataCamp's proven learning methodology is to practice all the information retained in courses. Using practice mode, learners can practice what they've learned with short challenges to test critical concepts. With over 3,400 practice questions, learners can practice their skills across various technologies and topics. Our mobile app is the perfect way to practice and learn on the go.

Apply

Once skills have been assessed, cultivated through courses, and sharpened through practice, learners are ready to apply their skills in a project-based environment. With DataCamp projects, learners can solve a variety of real-world R and Python data science projects. Learners can opt for guided projects, where they can follow step-by-step tasks and receive helpful feedback as they apply their newfound skills. They can also opt for unguided projects, which are open-ended, offering a variety of possible solutions and a live-code-along video to follow how an expert data scientist would approach a solution.

Track skills with skill matrix

Track the data skills your team has today and map a path to the skills they need tomorrow. Using the Skill Matrix, admin users can easily filter to identify individuals with the skills you need to take on specific projects or teams with low use or data skills gaps. They can then create and assign custom tracks to help bridge these gaps and report on skill development.

A robust training experience

We work with the largest brands in the world, including PayPal, Uber, HSBC, and Google, to help them transform their data skills. Our experienced team provides you with resources and guidance on everything from adoption best practices to SSO and LMS integrations—giving you the tools you need to upskill your team with confidence. Join over 7 million learners around the world. Close your talent gap. Visit datacamp.com.

More from DataCamp

Read our white papers and watch our own demand webinars