

Published in Insight

Anh Nguyen

Follow

Oct 2, 2018 · 7 min read · Listen

Save

# An Introduction to the Data Product Management Landscape



While companies have been building data products well before AI became the next big thing, the formalization of data product management positions is relatively recent. This is a direct function of data’s new-found role as a key product and competitive advantage. Modern u...e can reduce costs (e.g. optimizing data center operation), ... (e.g. driving cross-selling and

622

5

up-selling in e-commerce) or enable new breakthroughs (e.g. deep learning in autonomous vehicles).

Animal classification algorithms and Go-playing agents dominate the AI hype cycle, but algorithms are just one part of the entire data product ecosystem. In most business settings, the models might actually account for the least amount of impact. A large supporting ecosystem must be in place in order for data to flow through the veins of your organization:

1. Raw events and transactions need to be collected, stored and served
2. Data needs to be processed, discovered and shared with relevant teams.
3. Models need to be built, deployed and monitored in production.

And all of these undertakings need to produce concrete business results. How should an organization prioritize among thousands of potential directions?

At Insight, where we have helped thousands of Fellows transition into various roles in the data industry, we see a rise in industry demand for product managers who can tackle these prioritization and coordination challenges among data teams. This article aims to explain what product management looks like in the data space and why it is important.

### **Why Data Product Management?**

In small data teams without formal PMs, standard product responsibilities such as opportunity assessment, road-mapping and stakeholder management are likely performed by technical managers and individual contributors (ICs). This does not scale well for many reasons, the four main ones being:

1. Product work ends up accounting for all of the IC's time.
2. Not all ICs are well-equipped or willing to handle product work at scale.
3. Gaps between business units and technical teams widen.
4. Gaps between individual technical teams widen.

At this inflection point, there are two potential responses. The first approach is to break down work into projects that are self-contained enough for a single IC or small technical team to handle end-to-end, reducing the need for some type of central coordinators.

The second approach is to create a formal product management org that is responsible for maintaining source-of-truth roadmaps and coordinating different teams and ICs to execute. This is especially common for highly cross-functional products such as e-commerce and on-demand services.

If it's possible for a single IC to make adjustments to a product, immediately get objective feedback on how it's performing, and roll-back in worst-case scenarios without major ramifications, then the first approach is extremely powerful. While this might work for a free social network product, it's potentially catastrophic for a paid and operation-heavy product like on-demand services. Most companies at scale ultimately go with the second approach of having a product organization.

### **The State of Data Product Management Roles**

In the early days of the data revolution, orthogonal data skills like software engineering, statistics and modeling were rolled under the same umbrella of data science. These skills are rapidly being formalized into separate roles, such as data engineers, data scientists, research scientists and ML engineers.

Within product management, a similar trend is taking shape. Like their technical counterparts, we see the broad umbrella of Data PMs becoming further divided into sub-areas: infrastructure, analytics, applied ML / AI, discovery and standardization, and platform. These are not necessarily formal Data PM titles at the very moment. Rather, they reflect relatively distinct areas of data product work.

While each data use case requires a slightly different type of technical and domain understanding, which we touch on below, it is important to emphasize that generalist product management skills are still the most important drivers of success. 90% of what a Data PM does on a day-to-day basis will still be prioritization, communications, stakeholder management, design and specs.

### **Infrastructure**

At scale, individual product teams will have different use cases and data requirements. The natural tendency of these teams is to build out their own data

infrastructures in order to get started quickly. This tendency leads to duplicated work, data silos, and ultimately teams will run into similar data scalability issues.

The ultimate deliverable for an Infrastructure PM is a common data infrastructure that collects, stores and processes relevant data in performant manner to enable down use cases. This common infrastructure helps teams focus on using instead of collecting and storing raw data.

Infrastructure PMs' main KPIs are data availability, scalability and reliability. They are well-versed in data engineering technologies such as data ingestion, processing in batch and real-time, filesystem and delivery.

## **Analytics**

Decisions in the modern workplace are increasingly informed by data. Analytics needs to support a wide range of decisions, from strategy to product and ops, both offline and in real-time. While infrastructure PMs ensure that queries can be run efficiently on massive data sets (the *how*), analytics PMs focus on how to turn such raw data into actionable insights for decision makers like execs, PMs and ops team. In other cases, analytics PMs are also actively involved in defining key performance indicators and performing data exploration to help recommend business decisions.

Within the context of product building, an analytics PM is responsible for creating a mixture of self-service analytics, customized dashboards and reporting tools to help surface and share insights across an organization. Their stakeholders are diverse, from savvy data scientists to read-only consumers like execs.

The KPIs they are looking at are likely number of queries run, reports generated, etc. which indicate the ease for data users to extract the insights they need from raw data.

## **Applied ML / AI**

Certain products and features such as search, recommendation, fraud detection, etc. lend themselves naturally to ML / AI solutions. Applied ML PMs think about how data can be leveraged to improve an existing product (e.g. analyze chat logs to automate customer service routing) or how to design an entirely new experience altogether using advanced AI (e.g. filters for photo-sharing apps). Ultimately, they all work on directly improving key metrics for a user-facing feature.

PMs working on these features, though not always titled Data PM, generally have a strong understanding of the data science workflow and underlying machine learning models. They have strong intuition about leverage the power of ML while designing around its limitations to deliver a superior user experience compared to rule-based approaches.

## **Platforms**

As a company grows in size, the need for standardized frameworks becomes more apparent, particularly in experimentation and machine learning. The use cases in these two workflows are often very tightly integrated with the nature of the product itself, so few open-source solutions can truly meet the needs of everyone.

For this reason, individual data teams in large companies started with their own one-off systems, leading to duplicated work and slower time-to-market. The likes of Google, Facebook and Uber thus have embarked on platforming: common frameworks to help reduce the efforts spent on common tasks like tooling, deployment and monitoring.

These platforms aim to abstract away the need to manage data, deploy and monitor results, freeing data teams to focus instead on iterating around the models and the experimentations themselves. They also promote reusability by making common data and features accessible to all users of the platform.

Platform PMs start by proving out how the platform could be useful and convincing early adopters to give it a try. Once the platform has reached the inflection point, the role shifts to identifying high-ROI common denominators to build into the platform. They look at KPIs such as models or experimentations run on the platform, average time-to-market, etc.

## **Standardization and Discovery**

Standardization and discovery is yet another problem with growing data teams. As a company grows, the amount of data created by individual teams and people also grows exponentially. This rapid output of data creates a problem where there is no central place to see all the data that exist in an organization.

Without a structure to document, centralize and display metadata, the institutional knowledge of data sources is confined to the data owners. It becomes unclear what the data actually means, where they come from, how reliable they are, etc. Further,

any knowledge of these aspects of the data sources disappears when the employees most familiar with that data leave the team. Another common problem is that teams who use the same data often define similar-sounding metrics differently. For example, a team might define last-7-day as last 7 full days whereas another team might define it as last 168 hours.

A data standardization and discovery PM is responsible for ensuring that the entire organization becomes aware of the data that exist and use them in a consistently defined manner. A common product manifestation of this effort is a data catalog or a data portal that facilitates the discovery and definition of data / dashboards / metrics as well as identifies data owners who can be contacted for further conversations. A more advanced version of a data portal also makes computed metrics easily accessible and incorporated across different use cases (modeling, analytics).

## Final Words

The data product management landscape is still evolving and this is by no means an exhaustive overview of data product roles available in industry. Depending on the stage and organization structure of a company, the Data PM role could be a mix of these different responsibilities. Analytics might be a part of Infrastructure, and Standardization and Discovery might be a part of Platform. As an Applied ML PM, you might find yourself divesting resources to building out the infrastructure and deployment environments necessary to productionize your model.

Ultimately, these roles boil down to crafting a valuable data-powered user experience and removing all of the obstacles preventing a team from delivering that value.

*Interested in transitioning to a career in data? [Learn more about Insight](#), [apply today](#), or [sign up for updates](#).*

[About](#) [Help](#) [Terms](#) [Privacy](#)

Get the Medium app

