

Welcome to the newly launched Education Spotlight page!



[VIEW LISTINGS](#)

Data Science Central

A COMMUNITY FOR AI PRACTITIONERS

[Home](#) » [Business Topics](#) » [Data Strategist](#)

The Data Product ABCs – A Framework for Bringing Product Thinking to Data



[Juan Sequeda](#) | June 6, 2022 at 4:25 pm



A framework for data products

Let's be honest: The way we've been managing data for [the past 30 years](#) hasn't fundamentally changed.

Yes, the shift to the cloud and the Modern Data Stack is making the life of data engineers easier because you don't have to worry as much about infrastructure. Want data in a warehouse? Click, click, click... You got it! But if we zoom out, the foundations are the same: take data from source applications, integrate it into a target database then report, analyze, and fire up the ML/AI tools.

Bottom line — we've been dealing with data management from a purely technical perspective. And because of this, I argue we're driving ourselves insane; paraphrasing Einstein, we keep doing the same thing over and over again and expecting different results.

But there's hope. As a data community, we're starting to realize that we need a paradigm shift; data management can't just be about technology.



Welcome to the newly launched Education Spotlight page!

[VIEW LISTINGS](#)

Products.

Data Products: Product Thinking to Data

The phrase “Data Products” implies bringing [product thinking](#) to data management. This is already a common practice applied to software development; A product manager gathers feedback from customers, creates requirements and use cases from end-users, defines a roadmap, creates a plan, and manages releases.

The product team then executes on the plan by building in an agile fashion. They prioritize features to build, estimate the time involved, manage intake tickets, and test and release key functionality and improvements in accordance with the product roadmap. Product management teams care deeply about whether end users are getting value from the product, and they consider factors like feasibility, usability, viability, maintainability, reusability, and scalability in their decision-making.

Bringing this mindset to data is what leads to generating Data Products.



Welcome to the newly launched Education Spotlight page!

[VIEW LISTINGS](#)

more prevalent given the increasing popularity of data mesh.

In theory, treating data as a product sounds great, but I've found a handful of issues arise when it's time to put it into practice.

To overcome these issues, my colleagues and I at data.world have been working on a framework for Data Products which we call the [Data Product ABCs](#): **A**ccountability, **B**oundaries, **C**ontracts and **E**xpectations, **D**ownstream Consumers, and **E**xplicit Knowledge.

This framework provides insights into the types of questions data leaders should be asking when developing Data Products:

Accountability

Ownership and responsibility for the Data Product:

- Who is responsible for the Data Product?
- Who defines its requirements?
- Who fixes it when it breaks?
- Who is on call?
- Where does the liability sit?

Boundaries

Defining the Data Product and its interfaces:

- What is the Data Product?
- What isn't it?
- Where will it live?
- What are the inputs and outputs?
- What is the roadmap?
- How do you balance the roadmap against other organizational priorities and considerations?

Contracts and Expectations

Defining what makes the Data Product "good" (aka fit-for-purpose):

- What is the purpose of the Data Product?
- What are the use cases?
- How is data quality defined?
- What are the quality and maintainability details?



Welcome to the newly launched Education Spotlight page!

[VIEW LISTINGS](#)

- How is it being kept secure? Who can see it? Use it?
- What are the interfaces to it?

Downstream Consumers

Understanding the users and programs that depend on the Data Product:

- Who are the current consumers of the Data Product?
- Who are potential consumers?
- What are the use cases that have been considered?
- What is the value? How is it being measured?
- How will it evolve to provide more value for consumers over time?
- What is the consumption interface as it pertains to consumers' needs: well-defined tabular structure (CSV, XLS), API endpoint, SQL, SPARQL, GraphQL, Parquet, Graph, etc.

Explicit Knowledge

Making the meaning and semantics of the Data Product very clear:

- What is the meaning of the Data Product?
- What is the schema/ontology with corresponding definitions?
- What are the data constraints?
- How is it related to other data products?
- Where is the documentation with examples?

([A templated form can be downloaded at data.world.](#))

Note that this is not a complete and/or exhaustive list. Actually, I know it isn't. (What am I missing? I'd appreciate feedback from the community!)

Putting the ABCs framework to the test

We have been putting this data product framework to the test and have these learnings to share:

Code

In order to establish a standardized means of defining policies, contracts, and schemas, they must be computable, i.e. in code, and not just a pretty picture or text in a PDF. Well-established standards should be used for this purpose such as SQL, web standards-based on [RDF](#), [OWL](#), or [JSON Schema](#). Furthermore, select a language to define all the contracts and expectations: [Open Digital Rights Language \(ODRL\)](#), [Shapes and Constraint Language \(SHACL\)](#), [Data Quality Vocabulary](#),



Welcome to the newly launched Education Spotlight page!

[VIEW LISTINGS](#)

Modeling

Choose a standard architectural style such as Star Schemas, Data Vault, 3NF, etc.

Be explicit about contracts when you can. For example, a telephone number is defined as having a “country code,” “area code,” and “phone number”. Additionally, it has three types: “home,” “mobile”, and “work”. All telephone numbers must have a defined permission type, which can be “allow to call”, “allow to leave a voicemail”, etc. A “mobile phone number” is a telephone number that has a type of “mobile”. A mobile phone number must also have permission to send a SMS. The contracts can be used to enforce a schema, check for data quality, search, etc.

Create data products for the core business concepts (i.e. User and Activity).

Create data products for the Metrics (i.e. Real Active Customers) which are implemented as queries/calculations over existing data products of core business concepts (i.e. certain calculations of the activities that users produce).

Centralization and Decentralization

We are seeing that organizations are searching for the balance between centralization and decentralization that works best for their industry, size, and culture. Small- and medium-sized organizations start out being centralized, but over time they identify domains that want to take responsibility (and have the capabilities and resources to do so).

The trend is to push responsibility for the data to the domain team that knows and understands it best, and let a centralized entity (i.e. Center of Excellence) manage aspects that are for the greater good of an organization:

- The Accountability, Boundaries, and Downstream Consumers are aspects that should be managed by each domain. Contracts and Expectations, and Explicit Semantics should be *informed* by global standards in order to ensure agreement on semantics, syntax, contracts, policies, and access to data products.
- What’s most important, regardless of the interface, is the semantics, i.e the underlying logic. This includes keeping the Contracts and Expectations in place, and notifying producers and consumers if something goes wrong.
- A centralized entity can define the schemas for the core business concepts which can then be reused by data products generated by decentralized domains (e.g. telephone numbers). This way, different domains won’t define schemas for the same thing differently. Each domain can extend it if required.
- Of course, there are scenarios where imposition is necessary for security or regulatory purposes, as in GDPR.

People and Data Products



Welcome to the newly launched Education Spotlight page!

[VIEW LISTINGS](#)

We are now seeing the addition of two very important roles to new and/or existing data teams: Data Product Developers and Data Product Managers. These roles empower domain teams and help define the responsibility of data products:

- Data product managers understand your business's data ecosystem as a whole, facilitate the conversations between all data consumers and data producers — those who own and run operational systems — and define the data strategy, governance, and roadmap for the data products under their responsibility. They reduce the burden of data management across individual contributors, and they help data become more consistent, accessible, and accurate.
- Data product developers implement the data product within a domain with a key focus on knowledge and semantics. These may be data engineers, analytics engineers, or a combination.

An important aspect to consider when defining a data product team is measurement. It's crucial to explicitly understand who is consuming which data product and for what reason in order to understand if the data product team is being successful. Think of it in the context of ROI — Just like an e-commerce platform tracks what you click and when you buy, you can use data products to identify which assets are the most popular and drive the greatest usage.

Another social aspect to consider is if this level of decentralization will generate friction between data consumers. Should consumers be using the User data product of Domain A or Domain B? What happens if data is duplicated and we don't know which data to use? Are your teams wrestling with these questions and others?

If this type of friction is occurring, it's a sign of success! Friction implies energy, and it's an opportunity to embrace complexity. Now you can improve the communication between different domains, or maybe you realize you can globally enhance standards around schemas, contracts, constraints, etc. This is why the global standards must be developed in consensus by data product managers from each domain. But remember, do this in an iterative approach and focus where there is friction in order to avoid boiling the ocean.

Conversely, if there is no friction, everything is working well and you can leave it alone... or no one is using it, so there's no need to spend time on it.

Final Takeaway

The phrase "Data Products" implies bringing product thinking to data management. The Data Product ABC framework is a way to help understand what this looks like in reality: Accountability, Boundaries, Contracts and Expectations, Downstream Consumers, and Explicit Knowledge. Remember, this is not a complete and exhaustive list. Actually, I know it isn't. I would appreciate feedback from the community. Please reach out!

Juan Sequeda is the Principal Scientist at [data.world](#), and co-host of the weekly honest, no-BS data podcast, [Catalog and Cocktails](#).



Welcome to the newly launched Education Spotlight page!

[VIEW LISTINGS](#)

◀ PREVIOUS

[Comparing data fabrics, data meshes and knowledge graphs](#)

NEXT

[Building Value-driven Data Strategy and Economies of Learning – Part 1](#)

Related Content



[Economics of Ethics: Is Ethics Ultimately an Economics Conversation? Part I](#)

[Bill Schmarzo](#) | December 27, 2022 at 9:53 am



[The Technology Crutch and Crossing the Cultural Learning Chasm](#)

[Bill Schmarzo](#) | December 18, 2022 at 8:57 am



[Anatomy of a Use Case](#)

[Bill Schmarzo](#) | December 12, 2022 at 4:45 pm



[Digital Age Kool Kids Klub: Chief Data Officers and Chief Economists](#)

[Bill Schmarzo](#) | November 22, 2022 at 6:45 pm

[About Us](#) [Contact Us](#) [Partner with Us](#) [Advertise with Us](#) [Write for Us](#) [RSS](#)

[Legal](#) [Terms of Service](#) [Privacy Policy](#) [Do Not Sell My Personal Info](#) [Cookie Preferences](#)

© 2023 TechTarget, Inc.

