

Agenda

Data Domains, or how to Data as a Product

The right balance

Inspirations

What are data domains

How to structure around data domains

Challenges we met

Ekimetrics data projects roadmap

Ekimetrics data platform visions

Ekimetrics is a pioneering leader in data science and AI-powered solutions for sustainable business performance.

Who we are?

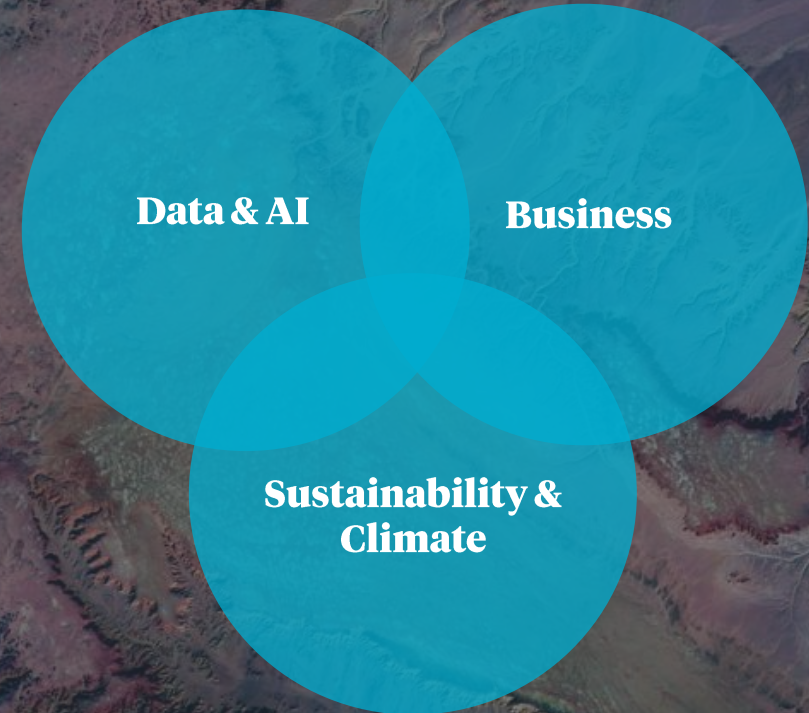


Ekimetrics

PARIS
LONDON
NEW YORK
HONG KONG

Technological expertise on Data Science and AI to multiply the positive impact

Proximity to the businesses and understanding of economic issues



Expertise on sustainability and climate issues and how data can accelerate transition

Data Domains, or how to Data as a Product

“ Shifting our thinking from data asset to data product shifts how we measure success from data volume to data user happiness. “

Zhamak Dehghani, Data Mesh

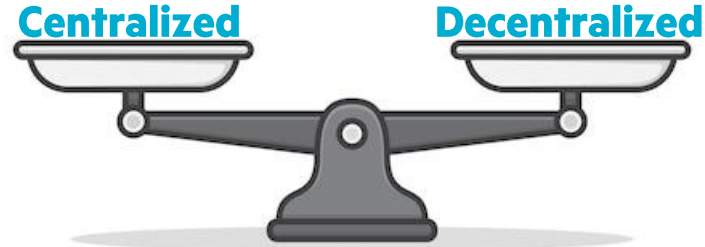
Whereas the first session was about the Why data as a product, this one is about the How.

The How to:

- organize data,
- organize teams around data,
- manage topics like quality, security, governance, costs, etc. in a fast and sustainable way,

And that's where we leverage Data Domains.

Looking for the right balance



Mostly centralized

- One rule or table can apply to all users
- Monolithic apps, monolithic databases, lakes, central governance team,

→ Monolithic approaches do not scale

Mostly decentralized

- Independent solutions by division, by topic
- Independent ERPs, datamarts, dashboards, data science applications,

→ Siloed approaches fail to deliver value

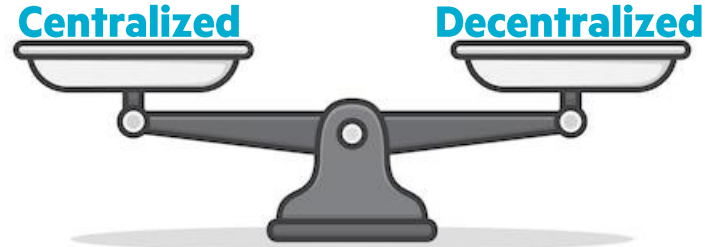
Federated governance around Data Domains

- Balance local autonomy inside data domains & global coordination between data domains

How?

1. When applicable, organize operational data into domains,
2. Organize analytical data into domains,
3. Structure roles and responsibilities within the domains

Looking for the right balance



Mostly centralized

- “Single source of truth” exist

Mostly decentralized

- “Single source of truth” projects fail (too fragile, too expensive to expand, etc.)

Federated governance around Data Domains

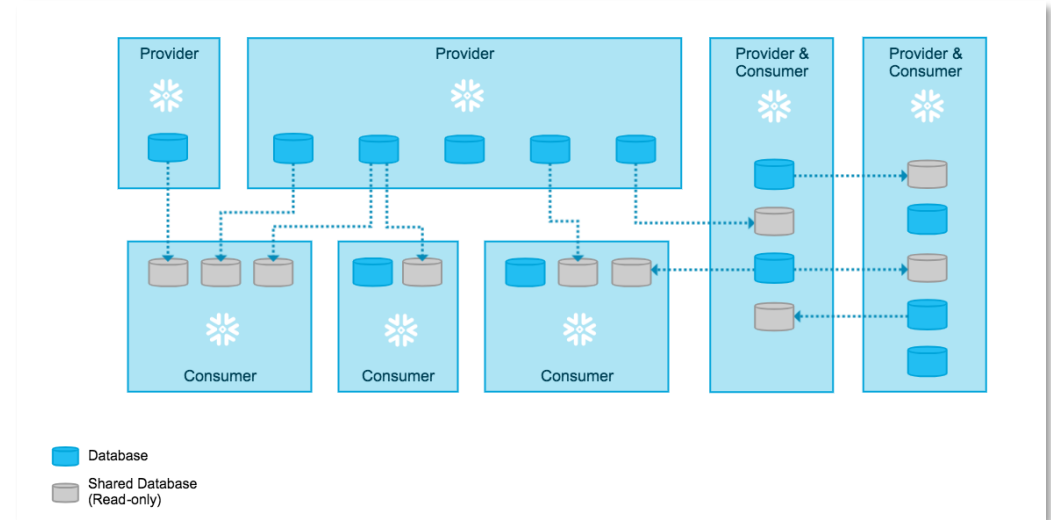
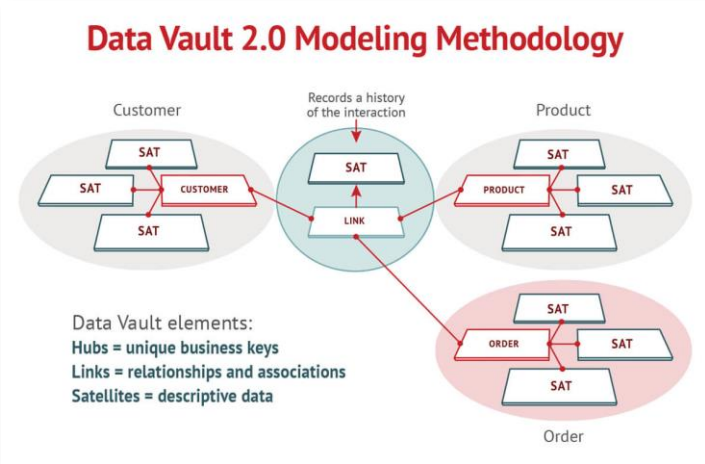
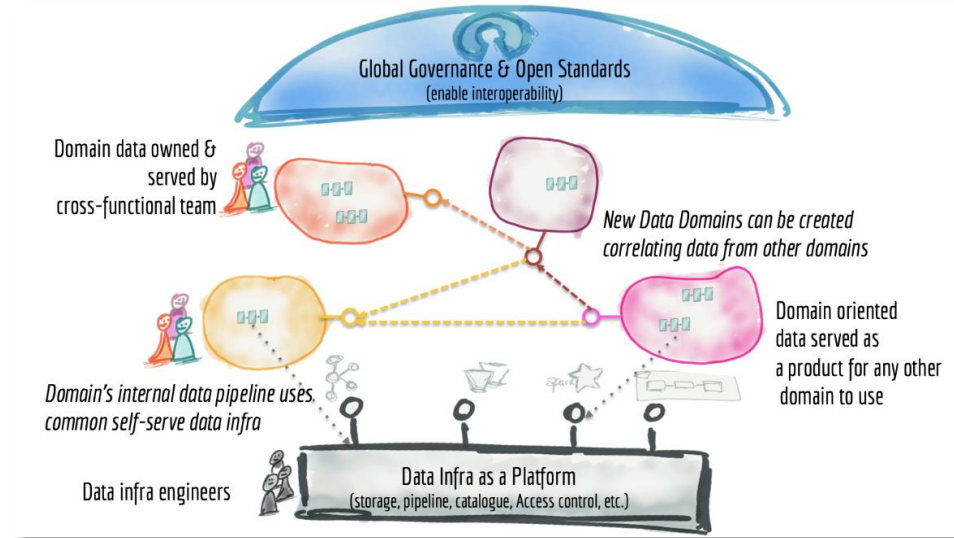
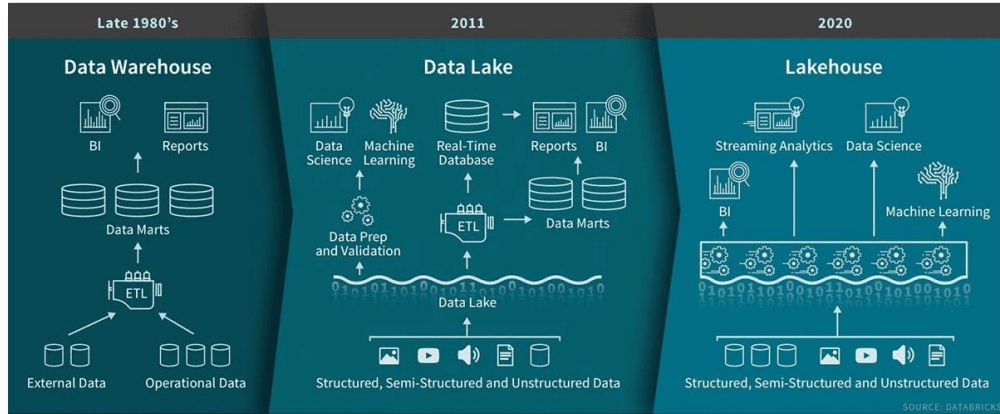
- “Single slice of truth” in each Data Domain, that don’t overlap

Bonus question:

- In which category do MDM (Master Data Management) platforms fit?

Tribute: “Single slice of truth” was coined by Zhamak Dehghani

Inspirations

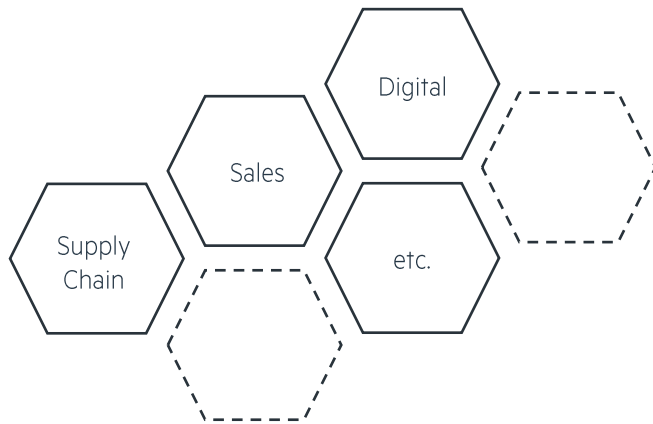


Sources:
<https://databricks.com/blog/2020/01/30/what-is-a-data-lakehouse.html>
<https://docs.snowflake.com/en/user-guide/data-sharing-intro.html>
<https://www.wherescape.com/solutions/project-types/data-vault-automation/>
<https://martinfowler.com/articles/data-monolith-to-mesh.html>

What are data domains

Data domains is a way to organize the content of a data lake to align it with the Business organization. It focuses on the responsibility of the Business teams over the data and the applications they own.

Multiple definitions of data domains correspond to different standpoints :



Standpoint	Definition of a Data Domain
Business & Finance	A set of data products, with costs and revenues based on the value of its use-cases
Architecture	A set of storage and compute resources, dedicated to a given business domain
Organization	A business team, in charge of data products, responsible of the quality of the datasets it shares with the rest of the organization
Interface	A black box, which interfaces are discoverable, consumable and trustable by the other data domains
Governance	A set of resources with a local governance, with both autonomy and compliance to global governance
Security	A bounded environment, with controlled and auditable access

How to structure around data domains

Data domains impact files or tables organization

An example structure of tables:

Domain	Organized by team that are responsible for data products
Subdomain	Organized by topic of the data.
Dataset	Organized so data is easy to consume

An example resulting table name : Finance . Costs . ProductsCosts

With the files folders in the lakehouse and the processing notebooks to follow identical structure.

Data domains impact security

An example of setup:

- For users of the data domain, data domain data is visible (if not stated otherwise through sensitivity groups)
- For users outside of the data domain, data domain data is invisible (unless specific access is granted)

How to structure around data domains

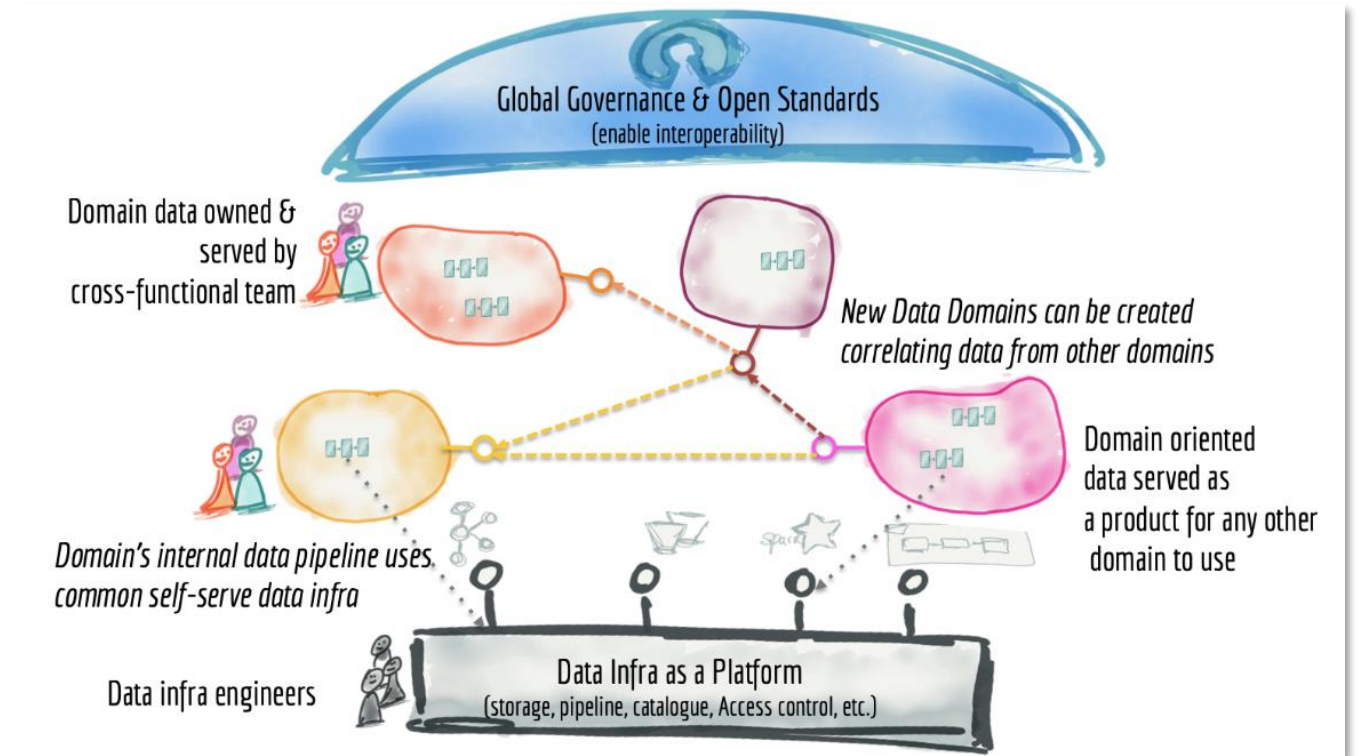
Data domains impact people organization

Data domains let you rethink the relationship between Data team, Data domains teams and IT team.

What we design here is very close to the diagram coming from Zhamak in her articles on Data Mesh.

Other team-related topics include :

- How squads should be rich enough to be autonomous,
- Moving from tickets to contracts,



Source: <https://martinfowler.com/articles/data-monolith-to-mesh.html>

Challenges we meet

Requires a strong CDO or a mature Head of Data

The understanding and sharing of the benefits of this approach requires some previous suffering with monolithic approaches.

Imposing Business-ownership of data assets is difficult.

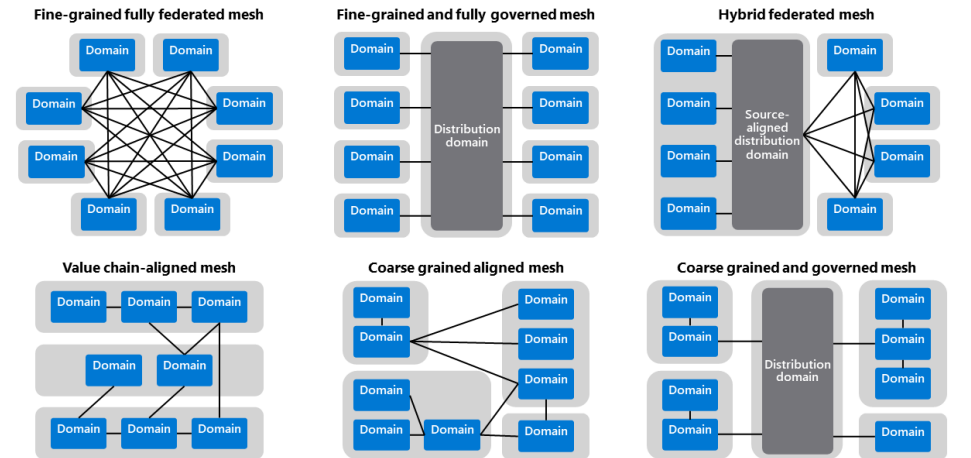
Need to balance roles and responsibilities between IT and Data teams

Especially in companies where Data does not “belong” to IT.

Need to balance between Business-owned and IT-owned data domains

As described by that article on the right that depicts nuances of mesh.

Governance Topologies : Different Approaches



Source: <https://towardsdatascience.com/data-mesh-topologies-and-domain-granularity-65290a4ebb90>

Ekimetrics data projects roadmap

(redacted slide)

Ekimetrics data platform vision

“ Shifting our thinking from data asset to data product shifts how we measure success from data volume to data user happiness. “

Zhamak Dehghani, Data Mesh

And because happiness is difficult to measure, Ekimetrics created the UUU score.

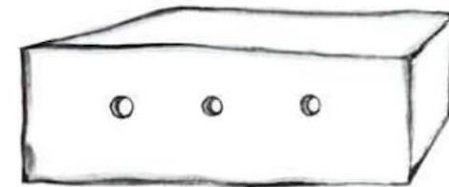
We see a growing appetite for data-mesh-inspired architectures.

As we deploy more and more of these solutions, we see the opportunity to:

- Automate some aspects of infra provisioning and data management,
- Accelerate development through contracts,
- Manage, from a business perspective, all governance, quality, security, etc.,

We are working on a platform to ease and accelerate Data Science.

Please come and talk if you are interested.



Contacts

Arnaud.lievin@ekimetrics.com

Baptiste.bon@yooi.com