MARTIN PETDER

BUILDING A BUSINESS DATA MODEL IN GLOBAL MANUFACTURING ORGANIZATION

WIZON

30.09.2025
Mövenpick Hotel, Tallinn

Tallinn Data Week 25

Building Data Models for an international manufacturing company

THE JOURNEY, LESSONS LEARNED, AND TECHNOLOGIES IMPLEMENTED FOR BUILDING BUSINESS DATA MODELS IN REPLIGEN

MARTIN PETDER, AIROMAR

Airomar: Technology Management Services

- **Process digitalization** and **IT/IS management** services for small and medium businesses
- Based on enterprise architecture principles, enabling **growth and efficiency** through a **product-based approach**
- Focus on manufacturing companies in a growth phase
- Consulting on data centric company operations
- Customers in 6 countries
- Goal-centric projects as well as multi-year service engagements

AIROMAR

TECHNOLOGY MANAGEMENT AS A SERVICE

airomar.com

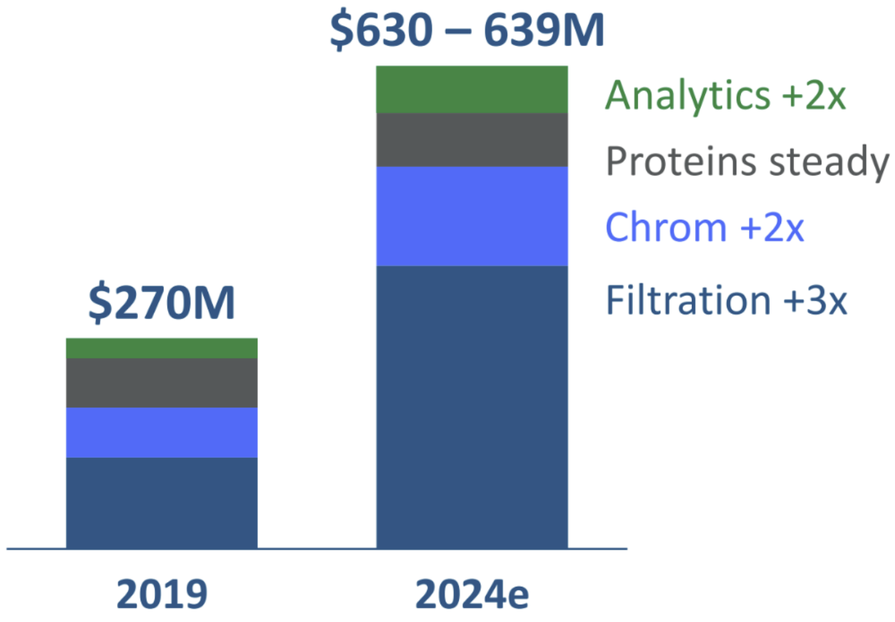
REPLIGEN Innovation Leader in Bioprocessing

- Supporting Pharma and CDMO's with a broad and differentiated portfolio of hardware & consumables used in their biological drug production
- Innovation engine … disrupting norms with fast-to-market products that enable yield gains and cost efficiencies
- Global manufacturing presence with security of supply
- ~65% clinical, 35% commercial
- Revenue majority mAb-based (~80%); strong and growing presence in new modalities (~20%)
- Operating at 41 locations in 14 countries

>1,700 employees

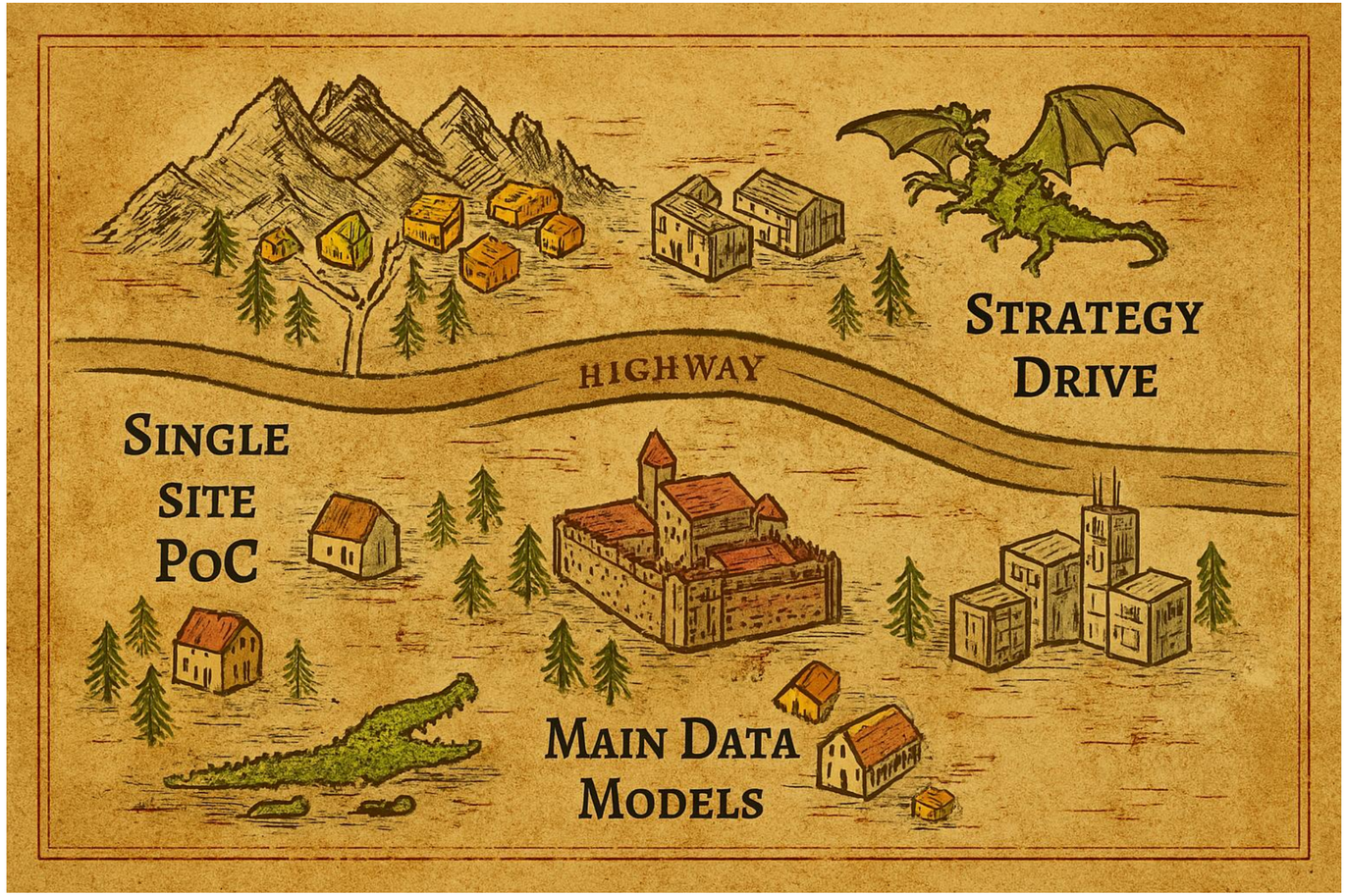
Performing Above Market

- **14** disruptive product launches … organic R&D and **14** acquisitions since 2014
- **19%** 5-year revenue CAGR

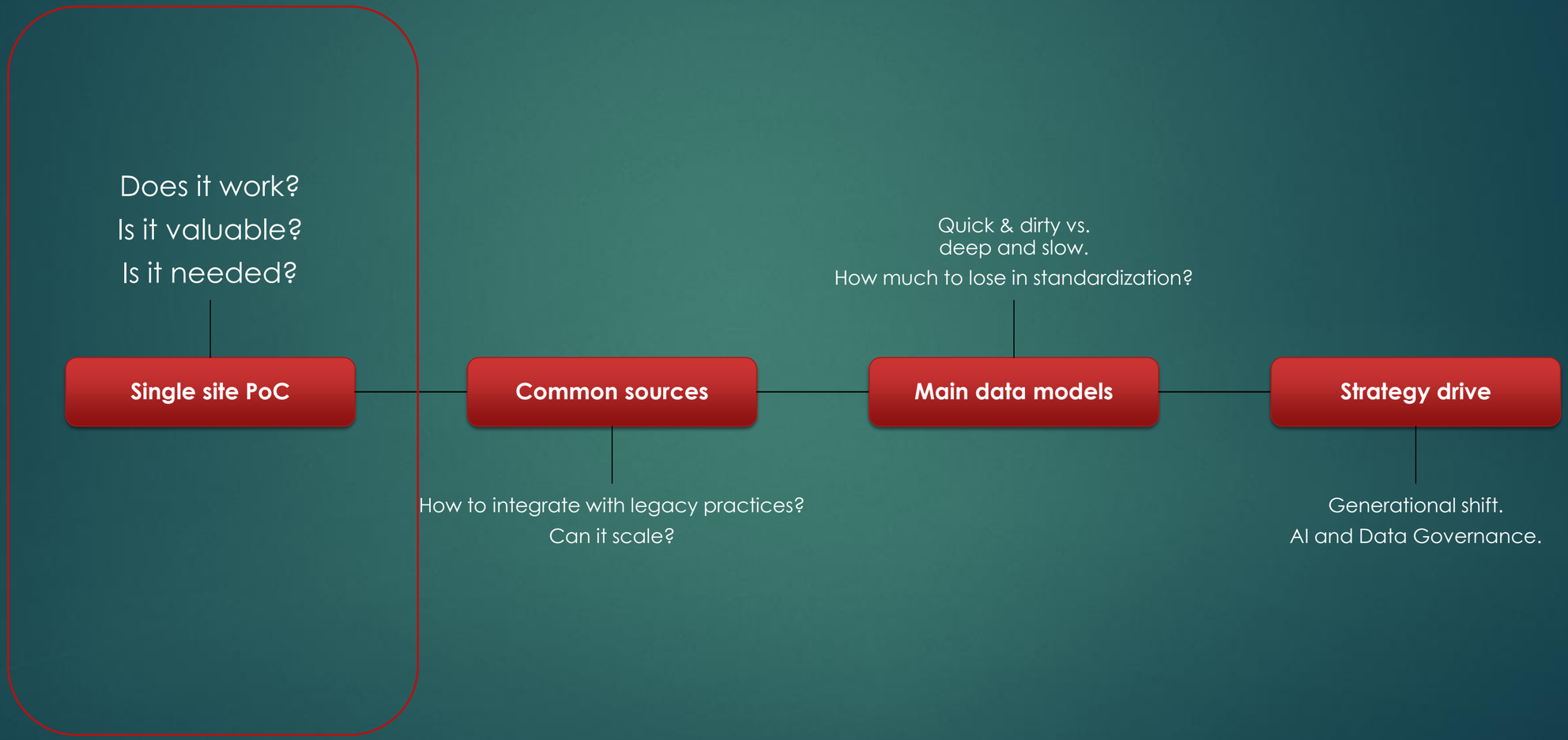


Data Journey

Almost greenfield data platform implementation with product team setup



Timeline


Does it work?
Is it valuable?
Is it needed?
Single site PoC
Common sources
How to integrate with legacy practices?
Can it scale?
Quick & dirty vs.
deep and slow.
How much to lose in standardization?
Main data models
Strategy drive
Generational shift.
AI and Data Governance.

Single site Proof of Concept

Accomplishments

Set up common data infrastructure

Established common principles

- ad-hoc from raw data
- stable reporting from the dimensional model
- some dimensional models and reports produced
- limited self-service via google sheets

Organization	Single site, two product lines, **single process domain**
Team	**Motivated business team** of process-centric managers 3-person data team + source system developers
Collaboration	**Open online collaboration** Weekly core team meetings
Technology	Snowflake storage **DBT modelling** Tableau worksheets Google sheets
Process	2-week iterations, **delivery as ready**, fixed burn rate
Governance	What governance? ☺

PoC outcomes

Positive

- Technology selection proved successful
- High value and (relatively) low cost
- First value within weeks
- Stable operations in 6 months
- Good collaboration with stakeholders

Challenging

- Highly dependent on stakeholder engagement - and it varies in time
- Blind spots in the business process identified.. and ignored
- Limitations in source data
  - Analog vs. digital processes
  - Data entry quality issues

Common concepts: Data Infrastructure

Build a test environment at the beginning

Will be hard to add later.

Yes, it's 2x infra cost – but it will be /4x development cost and allow for 10x better user satisfaction.

Build a repository and use it!

Build a CI/CD pipeline (and use it!)

Build a *fast-prototyping* solution – allowing to deliver in hours.

Keep tabs on what's developed – and move best prototypes into main solution before people get used to them ☺

Common concepts: Technology

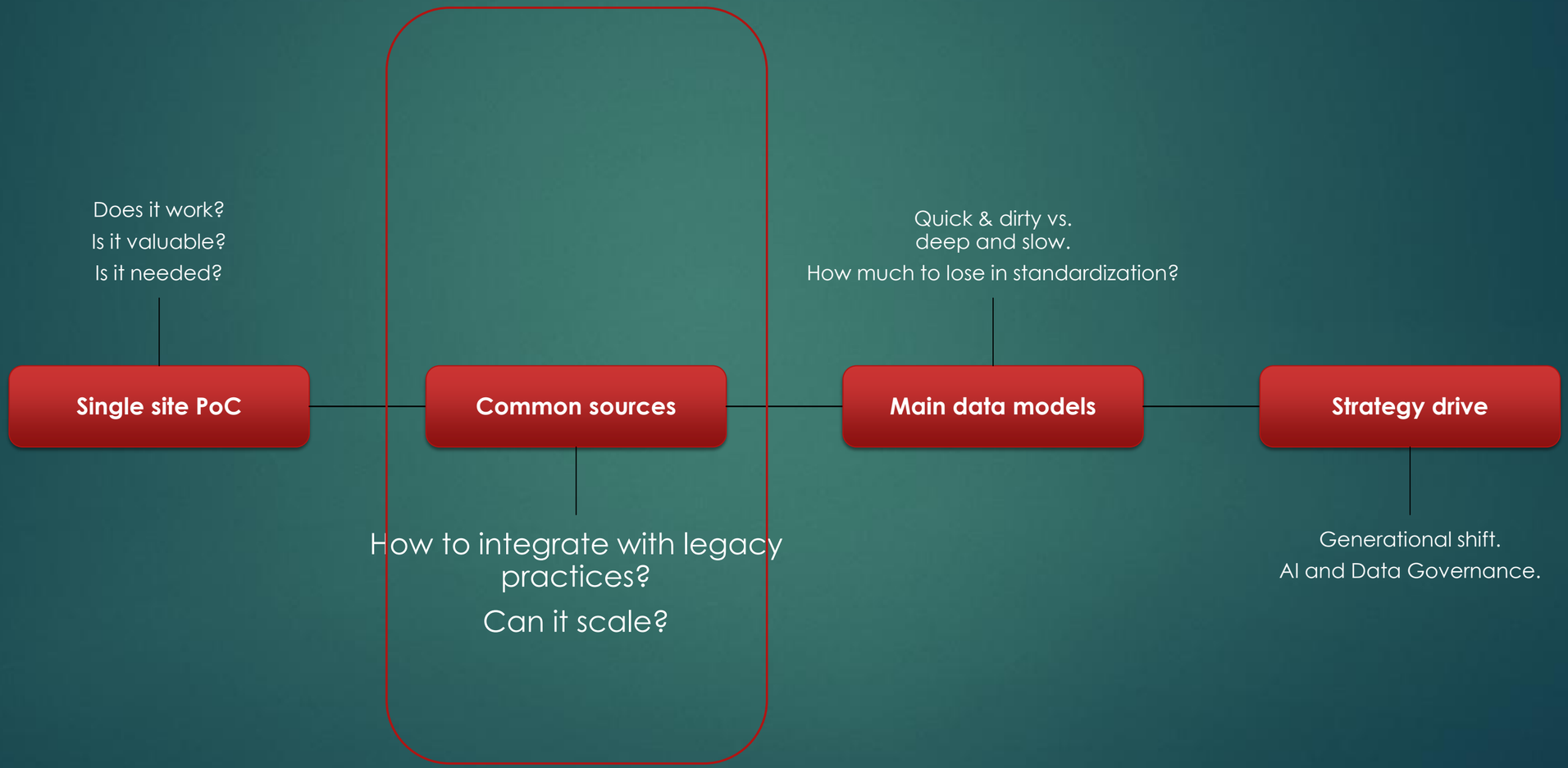
Commercial vs. DIY approach

- In the long run **DIY is cheaper**. But it might be **very long run**.
- **Commercial** allows to get there **fast**. Then to **learn** where "There" is.
- **Don't be afraid** to switch.

Languages and frameworks

- Stop using SQL when it **becomes unreadable**, especially **for new developers** ☺
- Be careful with visual modelling tools. Check their output, or you're in trouble.
- DBT is a good way to mix **simple SQL statements** with powerful data transformation.

Timeline


Does it work?
Is it valuable?
Is it needed?
Single site PoC
Common sources
How to integrate with legacy practices?
Can it scale?
Quick & dirty vs.
deep and slow.
How much to lose in standardization?
Main data models
Strategy drive
Generational shift.
AI and Data Governance.

Common sources

Accomplishments

ERP-to-ERP transition via data lake

Data management toolset in Google sheets w/AppScript

Introduction of data technologies to organization

Visibility into SAP processes

Data quality reporting

Organization	Single site, two product lines, **corporation core departments**
Team	Core team from PoC, but focused on transition **Master data teams** SAP development team
Collaboration	**Focus on meetings** No continuous collaboration
Technology	PoC platform Google sheets + AppScript **Outputs in Excel ☹**
Process	1-week iterations, **delivery as ready**, fixed burn rate
Governance	Hints of **master data requirements** emerging.

Common Sources outcomes

Positive

- Continuously **updating cleaned & curated** transition dataset
- Clear **understanding of data quality issues** (both at PoC site and beyond)
- Bringing **technology experimentation** into wider corporation

Challenging

- AppScript **performance troubles**
- **Cultural differences** in data
- Final transition via manually updated excel files
  - Automation **not quick enough**
  - Short time window
  - Known **quality issues in data** remained

Common concepts: Collaboration

From translation to a common language

- Data **engineering speaks foreign**, but **all must understand business**. The more, the better.
- Trust what you learn and learn how to trust.

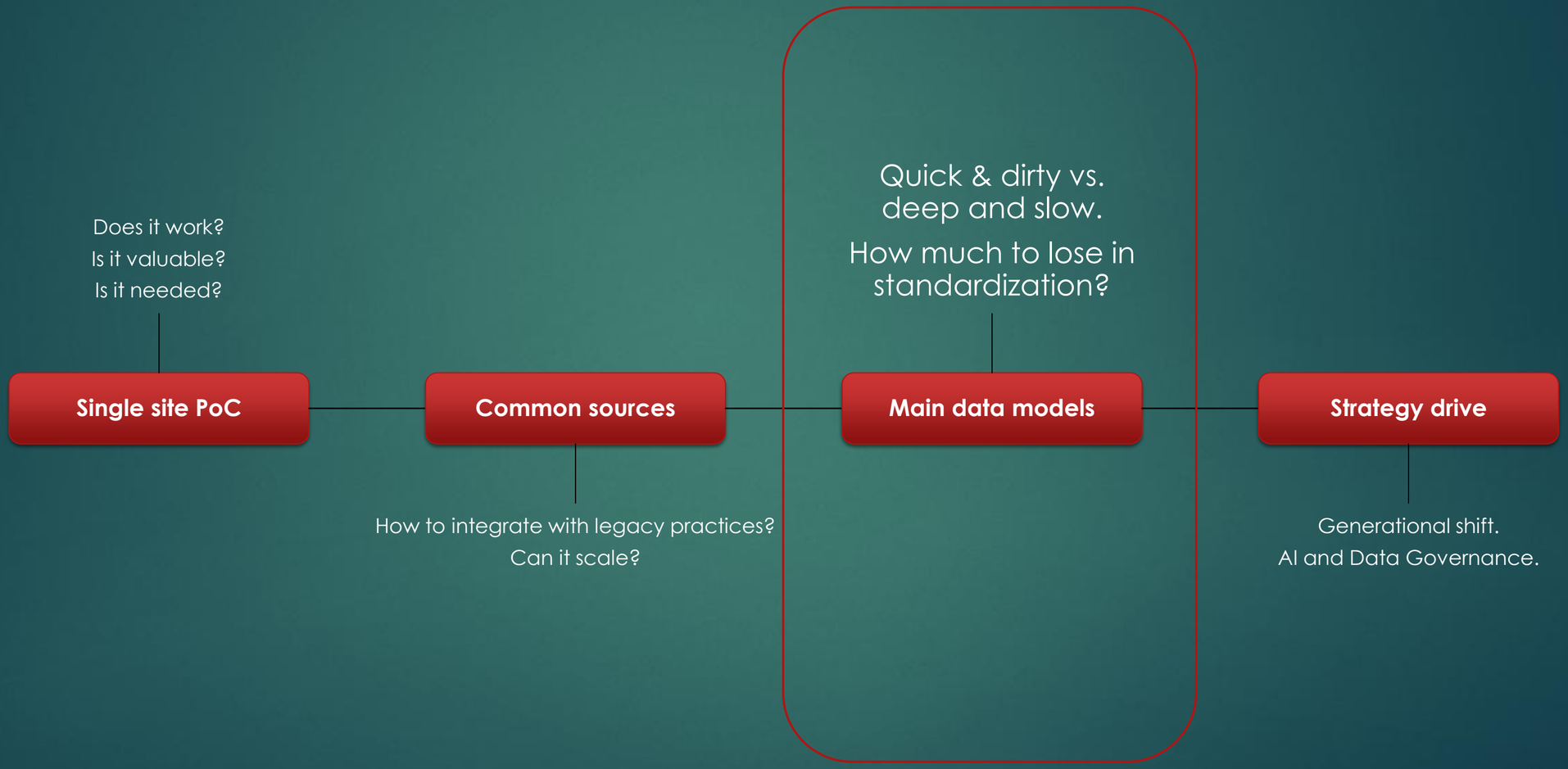
Common data models

- **Organizations are plural** in nature - standardization is nice, but very time consuming.
- Number of **supported corner cases define the value** of the model.

Don’t develop the organization, develop people

- No point to fight against strange processes – **help people in their needs** instead

Timeline


Does it work?
Is it valuable?
Is it needed?
Single site PoC
Common sources
How to integrate with legacy practices?
Can it scale?
Quick & dirty vs.
deep and slow.
How much to lose in
standardization?
Main data models
Strategy drive
Generational shift.
AI and Data Governance.

Main data models

Accomplishments

Critical business processes modelled

Frequent data updates (cost, not technology-limited)

Links to data quality management built in

Co-existence with other data consumption practices

Organization	The whole corporation, but **various levels of traction**
Team	Interview based approach Varying development team sizes per velocity requirements **Continuous core data team**
Collaboration	Core continuous collaboration **Focus-groups** when developing
Technology	No significant changes. **Switch to optimization** after every 3-4 deliveries
Process	**Dynamic approach**, based on the underlying need. **Continuous delivery**.
Governance	**Heated discussions** ☺

Main models outcomes

Positive

Standard modelling approach:

- Start from raw data & visual modelling for ad-hoc problems
- For important areas follow up with interviews & standard models
- Build dimensional model with linked quality tooling

Challenging

- Ad-hoc reporting proves very competitive ☺
- Building wider traction for Tableau reporting
- Balancing localization (not only language) and usability
- Getting ahead of the organization – and then behind when org changes
- Proper data management IS EXPENSIVE

Common concepts: Management

In the beginning it is all about Data

- Then analysts begin to **miss engineers** and management start to **miss responsibility**
- Then hackers arrive, promising speed, delivering chaos – and force improvements ☺

Splitting engineering and analytics early on

- Even if it's just 1 analyst and 1 engineer, but **set up for collaboration**

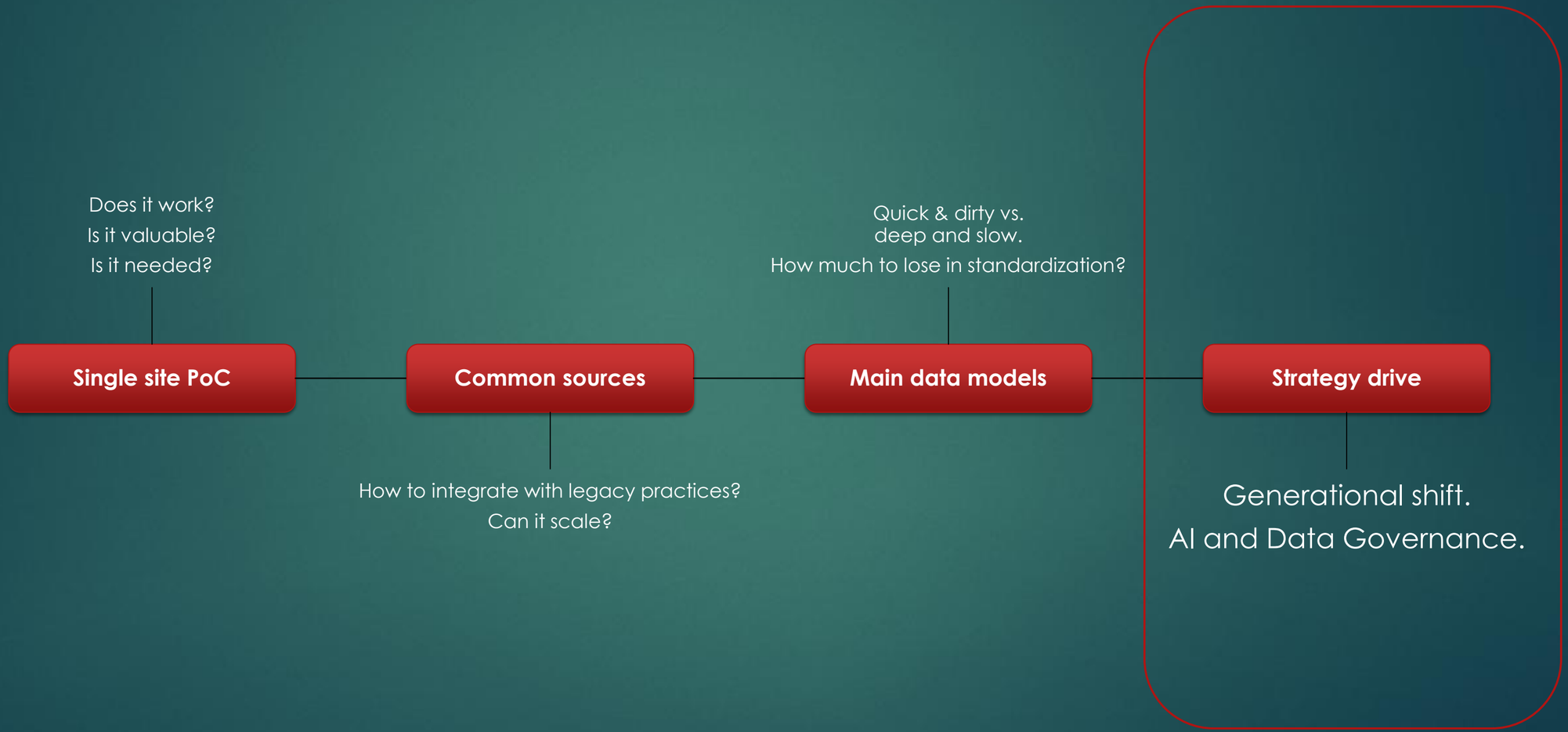
Make engineering talk to business

- The more developers **understand the need**, the quicker and **better it gets**
- Data models built in the ivory tower are just expensive decorations

Get business to take control

- **Eating your own practice** makes you resourceful
- Understanding the **needs down the line** add value to your output
- Yes, it might mean **technical people** are needed in the **business org**

Timeline


Does it work?
Is it valuable?
Is it needed?
Single site PoC
Common sources
How to integrate with legacy practices?
Can it scale?
Quick & dirty vs.
deep and slow.
How much to lose in standardization?
Main data models
Strategy drive
Generational shift.
AI and Data Governance.

REPLIGEN

WIZON

AIROMAR
TECHNOLOGY MANAGEMENT AS A SERVICE
airomar.com

Questions? ☺

CATCH UP DURING THE NEXT EVENTS OR USE THE QR CODE TO BOOK A MEETING.

GOLD SPONSORS

Bolt · Google Cloud · Sparkle · Microsoft

SILVER SPONSORS

WIZON · University of Tartu · SOLITA · SelectZero · veriff

BCS Koolitus · VALI IT · PROEKSPERT · TERASKY · dbt Labs

SUPPORTING PARTNERS

Supersimple · Tehnopol · Shroomwell Wellness Farming Innovation

Tallinn Data Week 25

Please give your feedback!

Help us be data-driven 😉
Scan the QR code to rate the event.

Tallinn Data Week 25