TEN DATA ARCHITECTURE STRATEGIES FOR IMPROVING EFFICIENCY, REDUCING COST AND RISK

ABOUT ME

- Data architecture, governance, and software delivery for 30 years
- Tier-1 banks & asset mgt., consulting firms, tech start-ups
- Lead and mentor architecture teams & data transformation programme mgt.

Jeremy Posner
LinkedIn:

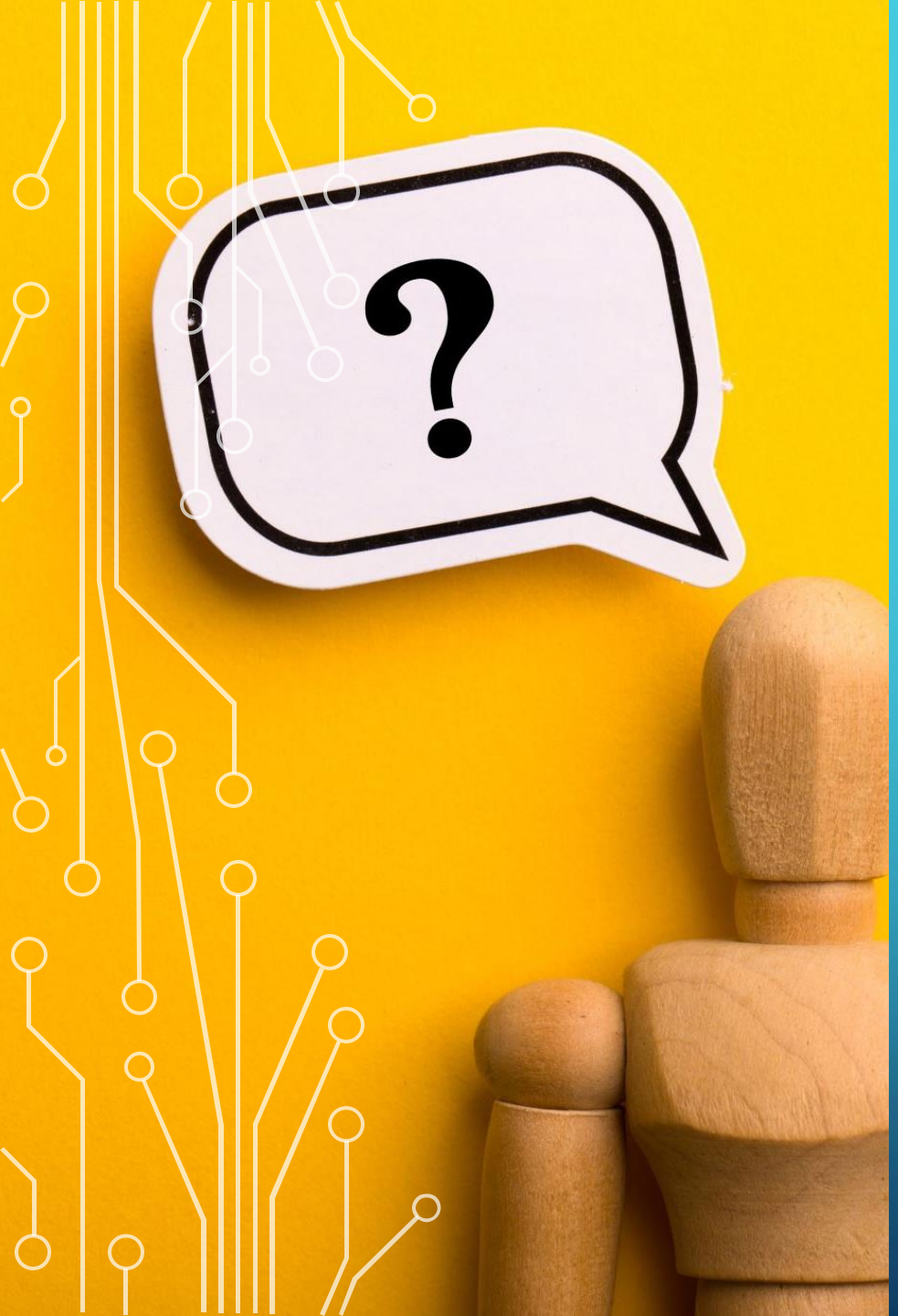
ABOUT YOU

You are probably here:

DATA GOVERNANCE (THE BUSINESS)

DATA ARCHITECT

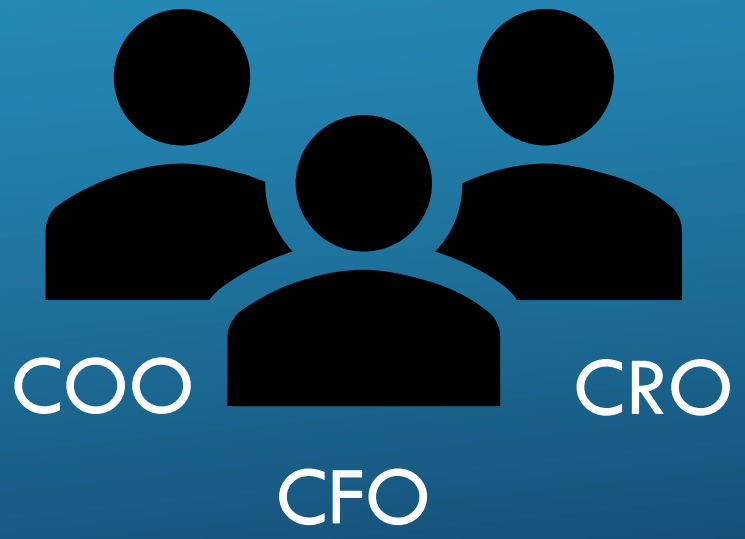
TECHNOLOGY DELIVERY (IT)



FOCUS OF TODAY

- NOT AI
- NOT DATA PLATFORM TECH
- NOT DATA MESH

EFFICIENCY/COST CONTROL RISK



ENDURING STRATEGIES FOR:

- Any Business Environment
- Any Technology Stack
- Any Data Architecture

1. STANDARDS & AUTOMATION

- Standardise everything you can.
- Standards => Automation => a Data Factory
- Don't reinvent the wheel. Use & extend standards where they exist

REAL LIFE EXAMPLE

- A Major German Bank had no standards in their APIs for how fields such as daycounts or securities were represented

- Different formats, different names for the same thing were used, usability was impaired, testing was tedious

- Defined standards and now all the APIs are using the same structures with shared validation.

- & Automated test harnesses for all the APIs

STANDARDS & AUTOMATION

STANDARDS & AUTOMATION - OUTCOMES

Efficiency

Reduce manual work, and speed delivery cycles. Standards allow delivery to do more with less.

Control

Standards ensure consistency, checks during build CI/CD process

Risk

Reduction in poor quality deliveries, remedial work, exceptions, user error

2. MODEL ENGINEERING

- Models are code / metadata, not just pictures
- Must drive technology artefacts (APIs & data structures)
- Selected model content should be pushed into other platforms, e.g. data catalogues and EA tools.

REAL LIFE EXAMPLE

- A US Investment Bank delivered its data model once per month

- Huge Model (1000s Entities), a change management problem

- Built a CI/CD mechanism to process the model, run checks vs. Data Standards, and deliver the model in minutes

- Can now deliver APIs and Database Design updates several times a day, with documentation, and change reports

- Also pushes outputs from the model to the Data Catalogue and to its EA Tools

MODEL ENGINEERING

MODEL ENGINEERING - OUTCOMES

Efficiency

Delivery of data structure changes in minutes.

Control

Keep a central model AND avoid it being a bottleneck

Risk

Higher quality data deliveries & greater consistency reduces DQ risks

3. EMBEDDED DATA GOVERNANCE

- Data Governance must not ‘chase the tail’ of delivery
- Embedding and automating a workflow involves business owners of data
- Linking Data Architecture artefacts to the Data Governance / Business users ultimately links Technology Delivery directly to the business.

REAL LIFE EXAMPLE

- A Global Investment Manager has a large Data Governance function & Data Stewards struggling to keep up with technology delivery function

- Lost control of the data in the Data Lake, risked becoming a swamp

- Connecting the Model-based outputs to the Data Catalogue and then adding workflow enabled the Data Stewards to review and approve data additions before they hit production

- Switching on Cloud Controls in their AWS environment enabled all data to be owned, classified.

EMBEDDED DATA GOVERNANCE

EMBEDDED DATA GOVERNANCE - OUTCOMES

Efficiency

Reduce need for DG admin functions & make Owners/Stewards more effective

Control

Governance built into the delivered data. Compliant with policies, audited, etc.

Risk

Reduced leakage of sensitive or incorrect/untrusted data

4. DATA PRODUCTS & CONTRACTS

- Products need product owners, but they also need standards and promotion
- A high-quality product has all the information needed to choose it, use it, support it.
- Data contracts are specifications for these products to allow wide re-use, facilitate quality and consistency

REAL LIFE EXAMPLE

- A Financial Data Provider provides 10,000+ datasets in many formats and access mechanisms, to hundreds of clients

- Huge duplication, because many of the data sets had repeated data, inconsistent names, and customers had demanded ‘specials’ – and was unmanageable

- Appointed several Product Owners who engaged clients to consolidate down the offerings.

- Now use a singular language from their model with standard product definitions and defined interface contracts, and they have cut their datasets by half

DATA PRODUCTS & CONTRACTS

DATA PRODUCTS & CONTRACTS - OUTCOMES

Efficiency

Reduced # of interfaces (complexity). Less code, more automation, less manual support

Control

Data Consistency, versioned, audited model-aligned

Risk

Less cyber attack surfaces, or chance of error from mis-use or mis-understanding

5. SYSTEM DATA AUTHORITY

- *Terms like: 'System of Record', 'Authorised Distributor', etc.*
- Important constructs - define what these terms are, and their meaning
- Record them and get them agreed, Now and Future
- Use when articulating the data architecture

REAL LIFE EXAMPLE

- A Global Bank has 2,000 Apps with ‘critical data’ and data quality issues which are under scrutiny by regulators with large fines looming

- Closing the DQ Gap requires them to know every application that masters any kind of critical data

- Defined System of Record at the Data Concept & Data Element level, and Authorised Distributor for downstream consumption.

- Data Quality Rules are now only applied at System of Record

- Non-Authorised Copies being remediated/switched off.

SYSTEM DATA AUTHORITY

SYSTEM DATA AUTHORITY - OUTCOMES

Efficiency

Reduce # of data masters & distributors. Remove redundant sources.

Control

Quality Rules applied at right places and available data through defined product interfaces

Risk

Reduction in likelihood of incorrect data due to bad sourcing or lack of controls

6. ENTERPRISE ARCHITECTURE DATA LENS

- EA typically focusses on Business Capabilities, Application Landscapes and Organisational / Technology Portfolio Optimisation
- Lacks the Data Lens, to answer key questions about data
- EA Tool > add Data Lens to identify potential rationalisation options

REAL LIFE EXAMPLE

- A Global Asset Manager uses an EA tool to manage business capabilities, mapping applications and organisations

- However it could not answer what data was where, how it got there, where it went, or which business process touched it.

- Approx 50 Data Concepts from the Data Model were synced into EA Tool. They now use these to describe data flows, link to business process and define information lifecycle

- Can now articulate where Market data comes into the company and how it flows through the applications, allowing consolidation programmes to reduce spend

ENTERPRISE ARCHITECTURE DATA LENS

ENTERPRISE ARCHITECTURE DATA LENS - OUTCOMES

Efficiency

Find sub-optimal patterns & eliminate waste. Design a leaner new world

Control

Describe the whole landscape, including data viewpoint. Ensure new projects align to strategy & roadmaps.

Risk

Reduce both failed project risk and duplicative project risk through EA assurance

7. DATA LINEAGE STORIES

- Be Selective - don’t try to boil the ocean
- Choose High Impact areas
- Use Lineage to focus on business pain and excessive cost.

DATA LINEAGE STORIES

REAL LIFE EXAMPLE

- A European Bank has 15 legacy Data Warehouses and plans to consolidate them into a new Cloud Data Lake

- Funding this work requires understanding the complex flow of data to provide estimates for delivery, at least 5 years.

- Use a formula for estimation, based on the number of data flows, and the content of each, and technology involved

- To gain funding, they told the data lineage story using their EA Tool, showing key data concepts moving between each source system and each Data Warehouse.

- The new Cloud Data Lake will require 200 data flows instead of the current 1000+. Savings estimated at EUR 50m annually.

DATA LINEAGE STORIES - OUTCOMES

Efficiency

Gives the visual ammo needed to show inefficiencies for cost reduction

Control

Improved data flow paths will improve Data Quality

Risk

Remove Risks of Incorrect data sourcing, Data Quality issues and Cyber attack surface

8. SDLC CHECKPOINTS

- SDLC is often missing the data lens
- Architectural input early
- Gates during development
- Checks as part of release management check-list

SDLC CHECKPOINTS

REAL LIFE EXAMPLE

- A London-based Alternative Investment Manager has a SDLC but no references to good data delivery

- Engineers do their own thing, and the single Data Architect struggled to keep tabs.

- The Data Architect worked with CTO office to add 3 steps in the SDLC where projects required check-in with Data Architecture and/or Data Governance: 'Planning,' 'Development', and 'Pre-Release'.

- These Data checkpoints educate Engineers, point them to standards, tools and reusable components, as well as ensure higher quality data delivery that complies with internal policies

SDLC CHECKPOINTS - OUTCOMES

Efficiency

Better planning of resources, effective reuse, more supportable solutions, less rework

Control

Enables standard controls to be injected into all projects, and adherence monitored

Risk

Lower risk of poor quality deliveries, remediation work, data quality issues.

9. DATA ARCHITECTURE DEBT MANAGEMENT

- Things go into production that shouldn't
- Every exception to an agreed path ultimately costs
- Track, prioritise and manage data debt
- Risk management is an essential tool if you want to clean up

DATA ARCHITECTURE DEBT MANAGEMENT

REAL LIFE EXAMPLE

- A UK Retail Bank has standards and guidelines for data, runs ARB and has an SDLC, but still delivery projects were allowed to go into production to hit tight business deadlines.

- The EA Team now creates and maintains an Architecture Debt backlog, including 'Data Debt' - This is referred to whenever a new project or programme increment is planned.

- The priorities from previous delivery's Debt Register are reviewed, scored and estimated as part of the next increment

- Any failures to remediate are escalated at firmwide Risk meetings

DATA ARCHITECTURE DEBT MANAGEMENT - OUTCOMES

Efficiency

Exceptions from agreed strategy and direction are costly to maintain, difficult to manage and reduce capacity/speed

Control

Staying aligned to architecture strategy keeps in compliance with standards and control frameworks

Risk

Management of Data Risks can be tracked, measured and escalated to higher levels.

10. EXPOSE KNOWLEDGE

- The ‘data stack’ carries huge knowledge, but the metadata is fragmented and not easily queryable.
- Integrate and share to the Organisation, for multiple purposes, e.g. Operational Resilience, or used for better GenAI
- Expose this metadata in standards-based forms, for interoperability

REAL LIFE EXAMPLE

- A US Investment Bank has over 20 tools and repositories of data architecture and data governance information.

- The bank requires all the data related information to be integrated and queryable by Operational Resilience, Legal, Compliance and Cyber Security teams.

- It creates a connected set of data, defined using standard ontologies, in its 'Enterprise Knowledge Graph' - a federated set of stores queryable using open standards (RDF, SPARQL)

- Each downstream consumer now has access to the full set of connected and integrated data, solving some of its most difficult data challenges

EXPOSE KNOWLEDGE

EXPOSE KNOWLEDGE - OUTCOMES

Efficiency

Integrated, extensible and reusable metadata available via open standards

Control

Standard ontologies for describing the environment and configuration/inventory

Risk

More granular data available for new Regs such as DORA

10 STRATEGIES – A SUMMARY

Standards & Automation

Model Engineering

Embedded Data Governance

Data Products & Contracts

System Data Authority

Enterprise Architecture Data Lens

Data Lineage Stories

SDLC Checkpoints

Data Arch. Debt Management

Expose Knowledge

THANK YOU