



# Foreword



In my role, I see first-hand how the conversation around data has evolved across the lending sector. What was once considered a supporting function is now central to how organisations manage risk, serve customers and make decisions.

At the same time, expectations have increased. Regulators are demanding more from firms, leadership teams want clearer and more timely insights, while interest is growing in how data can support areas such as automation and advanced analytics. For many organisations, this creates both momentum and challenge - particularly when trying to balance ambition with the realities of existing systems and processes.

This paper reflects a number of those challenges in a practical way. It brings together perspectives from across the market and explores how the different components of data modernisation fit together, from strategy through to architecture, engineering, governance and analytics.

One of the points that resonates most is that there is rarely a single “right” answer. Progress tends to come from making pragmatic decisions, improving incrementally and being clear about priorities at each stage of the journey.

It also highlights that technology alone is not enough. The organisations making meaningful progress are those that align ownership, operating models and culture alongside their platforms.

For those navigating their own data modernisation journey, I hope this provides a useful perspective and helps bring some clarity to what can often feel like a complex and evolving space.



**Sam Bridges-Sparkes**

Head of BI Analytics  
& Strategy

Shawbrook Bank Limited



# Table of Contents

## Data Modernisation Whitepaper ..... 1

Introduction.....	4
Purpose.....	4
Market-wide perspective.....	5

## Data Strategy ..... 7

Introduction.....	7
The regulatory backdrop.....	8
Why AI changes the stakes.....	8
What we see happening at institutions.....	8
What a good strategy looks like.....	9
Building your strategy in practice.....	11
The relationship to what follows.....	11

## Data Architecture ..... 12

Introduction.....	12
What's currently happening at institutions?.....	13
What does success look like?.....	13
Let us explore the common approaches to modern data architecture today.....	13
What's the north star?.....	17
How to implement a modern data architecture.....	18

## Data Engineering ..... 20

Introduction.....	20
From Tactical Extracts to Controlled Pipelines.....	21
Engineering is Pragmatic.....	22
The risks of not taking data engineering seriously.....	22
How to implement data engineering.....	24

## Data Governance, Management & People ..... 26

Introduction.....	26
Governance begins with ownership.....	26
Maintaining Governance with Changing Operating Model.....	28
Governance During Modernisation.....	28
Importance of Transparency and Hygiene.....	30
Governance in a Hybrid and Distributed Estate.....	31
Data Culture and People.....	32

## Data Analytics & AI ..... 33

Introduction.....	33
Data Analytics Today.....	33
Taking Analytics Further.....	34
Should you enable self-serve?.....	37
How to uplift your Analytics.....	37

## Looking Ahead ..... 38

Your Data Modernisation Journey.....	38
What comes next?.....	39
The Role of Strong Data Foundations.....	40



# Data Modernisation Whitepaper

## Introduction

This whitepaper delves into the modernisation journey of establishing a robust and scalable data capability as a regulated lender. The contents are tailored to banks, building societies and specialist lenders and features actionable insights from market participants.

Across strategy, architecture, engineering, governance and analytics, the message is consistent: strong data foundations are now a prerequisite for a competitive advantage, regulatory confidence and operational resilience. The challenge is rarely understanding what good looks like, instead the difficulty lies in pragmatic decision-making, investment, managing change while demonstrating measurable business value.

## Market-wide perspective

Across the sector, ambition has risen sharply because of AI, but many firms still approach data through the old lens of cost justification rather than strategic enablement. That tension, along with the budget constraints, helps explain why so many institutions have partial modernisation rather than end-to-end transformation.

For this report, Woodhurst spoke to six regulated lenders at different points in the data modernisation journey.

Two thirds of the engaged lenders are mid-journey; they have implemented basic infrastructure and a formal strategy, with engaged leadership but have not yet converted

to a fully operational data capability, with governance frameworks still a work-in-progress.

The other lenders we spoke to are further along the modernisation journey, with an advanced and operationally embedded modern architecture. One lender is re-platforming, expanding access through tools and self-service, aligning analysts to domains and actively tracking data quality. Over the past few years, they have moved from legacy reporting and on-premise warehousing through Synapse to Microsoft Fabric, establishing a stable base. Notably, their operating model is modernised alongside the platforms, with a governance committee, strategy, data owner hierarchy, steward structures and a deliberate attempt to bring the business with them.

So, why are lenders investing in data? From the obvious to the lesser known, we have outlined the **drivers for modernisation**:

- **Compliance benefits:** Modern tooling provides metadata-driven lineage and provenance, enabling better tracking and monitoring compared to legacy black-box processes. This can enable benefits such as automated regulatory reporting, real-time AML/ fraud monitoring or provide a clear view of an organisation's GDPR position, supporting compliance efforts.
- **Enhanced security:** Modern environments offer granular access control down to individual columns for PII data and GDPR compliance, reducing insider risk as staff cannot see data that is not directly relevant to them, reducing the attack surface as less visible data means there is less to steal or misuse and enabling greater traceability as access to sensitive fields is logged and auditable.
- **Reducing change complexity:** Consolidated, system-agnostic architecture means replacing a core system requires building one set of interfaces instead of six, dramatically reducing testing and downstream reporting work (from potential year-long projects to three months).
- **Future proofing:** Ensuring data acts as a strategic enabler and not a constraint, allowing organisations to fully realise the value of new initiatives.

Many transformation efforts fail due to underlying data limitations; with the rise of AI, these shortcomings are amplified, further hindering progress. Modernising data capabilities is therefore critical to unlocking scalable, compliant and sustainable change for regulated lenders.

- **Cost effectiveness:** Consolidating tools, platforms and skills across the organisation, reducing the number of different platform fees.
- **Career development:** Unified tooling creates transferable skills, enabling colleagues to progress their careers and move within the organisation more easily. With a unified tech stack across an organisation, barriers for moving into new roles across the organisation are minimal compared to organisations with various software/platforms.

## Common Themes

Most lenders are modernising in waves: first stabilising reporting and infrastructure with a solid foundation in the technology stack, then improving access and ownership governance and only then trying to unlock analytics, AI, or self-service.

### Wave One

Stabilising reporting & infrastructure



### Wave Two

Improving access & ownership governance



### Wave Three

Unlocking analytics, AI or self-service



**The more mature firms are not necessarily the ones with the most advanced technology stack; instead, they have aligned leadership, clear data ownership and treat data as an operational and strategic capability.**

Our conversations with regulated lenders at various stages of maturity have uncovered common themes:

- **Data modernisation has moved from an efficiency-focused agenda to strategic necessity:** Historically, the conversation was cost, automation and reporting. AI has raised the stakes and turned data quality and accessibility into a strategic focus.
- **Leadership attitude is a determining success factor on the modernisation journey:** Where leadership treats data as a strategic lever, modernisation is quicker and more effective. Executive backing has reduced friction and made data quality and governance easier to land as an enterprise-wide priority, not just technology concerns. Where leadership treats it as a cost centre, progress is slower and more fragmented, with increased hurdles around change management and adoption.
- **Technical progress often outpaces organisational maturity:** Platform modernisation is tangible and adoptable, while ownership, stewardship and business fluency are harder to operationalise. Setting up the organisational structure and managing the people-side of change takes time and means taking your colleagues on the journey to get the buy-in required for success.
- **The market is converging on hybrid operating models:** Regulated lenders are tailoring solutions to their specific business, with governed decentralisation: central standards with domain-aligned capability. This allows each business area to focus on their unique objectives, while leveraging centralised, organisation-wide data that provides the full picture.
- **Governance is becoming more pragmatic:** Governance is being balanced, so that it is applied proportionately without creating so much red tape that delivery slows.
- **Migration is more of a data problem than a technology problem:** Core replacement and platform change expose hidden data dependencies, undocumented processes and semantic inconsistency. Uncovering and resolving these decade-old issues impact delivery timelines with amplified downstream impacts but usually become known in the discovery phase once timelines have already been determined. An example we've seen is the difference in account numbers formats, that flow into almost every report and overall customer experience, down to the passbooks. The solution: use discovery to uncover data problems, understand realistic timelines and do not make assumptions ahead of time.

**Think ahead of the implementation, in terms of the data you want. Data requirements are typically engaged too late in project lifecycle, often a few months before go-live. By then, system configuration is already complete, making changes impractical without disproportionate cost or risk...Without capturing data on how our customers utilise our services, you risk building digital products that don't meet member needs.**

Steven Forth, CFO,  
Darlington Building Society

The strongest common message is that AI has increased the urgency to modernise; firms increasingly recognise that advanced analytics, automation and customer-facing intelligence depend on foundations that are still uneven across the market. Executive sponsorship, pragmatic governance and operating models that bring data capability closer to the business appear to be the main factors separating progress from stagnation.



Our data program is a key part of the bank's infrastructure investment... The need to modernise and generate efficiencies to fuel growth is one of the big strategic pushes for our organisation. Executives want to do it sustainably and to do it well - If you draw any line from that strategy and how you do that, you will eventually hit data. This highlights that unrestricted access to good quality data is imperative.

David Holton, CTO  
Cambridge & Counties Bank



# Data Strategy

## Introduction

A data strategy is the document that ties everything else in this whitepaper together. It defines what an institution wants to achieve through data, which capabilities need to be built or improved to get there and in what order. Without it, decisions about architecture, engineering, governance and analytics tend to be made in isolation - technically sound but pointing in different directions.

This section covers what a good data strategy looks like for a regulated lender, why most institutions do not have one that functions in practice and how to build one that does.

In 2023, the Basel Committee on Banking Supervision published its sixth progress report on BCBS 239 - the global standard for risk data aggregation and reporting that has been in force since 2013. After a decade of effort, only two of the thirty-one assessed global systemically

important banks were fully compliant with all principles. Not a single principle had reached full compliance across all assessed institutions. And compliance had actually fallen by 2.6 per cent since the prior review.

These are the world's largest, best-resourced financial institutions. If they cannot get data right after ten years of trying, the problem must be structural.

For UK regulated lenders, the implication is uncomfortable. The instinct to treat data as a technical problem - something fixed by the right platform, the right hire, or the right vendor - has not worked for the biggest banks in the world and it will not work here. Data capability is built through sustained strategic commitment, consistently resourced and governed over time. Without that, initiatives disconnect and the gap between ambition and delivery gets wider.



This section sets out why that commitment needs to be explicit – framed as a strategy – before decisions about architecture, tooling, or governance are made.

## The regulatory backdrop

For most of the past decade, data governance sat in the category of sensible but discretionary. The regulators have been trying to change that.

Regulators have made data quality a first-order supervisory concern. In 2025, the PRA’s “Dear CEO...” letter to lenders explicitly identified poor data quality as a recurring issue, affecting firms’ ability to manage risks and submit accurate regulatory returns. The FCA, in a November 2025 review of prudential regulatory reporting, ran over 320,000 data quality tests across nearly 3,800 investment firms. Four in ten failed. The regulator was direct: accurate, timely and consistent data is not optional.

Consumer Duty has added further pressure. The FCA has identified over 200 specific data points firms are expected to monitor and act on to demonstrate they are delivering good outcomes for customers. The evidence the FCA expects firms to provide is fundamentally a data challenge before it is anything else.

The consequences of getting it wrong are well-documented. TSB’s 2018 core banking migration – which locked 1.9 million customers out of their accounts – resulted in £318 million in migration costs, £247 million in remediation, £49 million in regulatory fines and a personal fine of over £80,000 for the former Chief Information Officer. The root cause was not a technology failure in the conventional sense. It was insufficient understanding of the firm’s own data estate and inadequate testing of how that data would behave in a new environment.

**What was once considered good practice is now supervisory expectation. Institutions treating it as the former are increasingly being held to account as though it were the latter.**

## Why AI changes the stakes

At the same time, the strategic case for getting data right has become harder to ignore. Generative AI and advanced machine learning are being used everywhere. They are being applied to credit decisioning,

arrears propensity, customer vulnerability detection, regulatory reporting and a plethora of emerging use cases across the sector.

In February 2025, Gartner forecast that 60 per cent of AI projects would be abandoned through 2026 – not because the AI does not work, but because the underlying data isn’t ready to support it. Separately, between 70 and 85 per cent of AI project failures trace back to poor data quality and fragile data pipelines rather than to any problem with the models themselves.

The institutions best placed to benefit from AI are those that have already consolidated their data into a well-governed, documented estate, with clear quality standards and lineage. Those that have not are starting from further back and the gap tends to widen rather than close. Deploying AI on top of a fragile data foundation can actually make things worse, rather than better. For example, implementing AI driven credit decisioning on flawed data, can lead to a misunderstanding of credit risk and lenders facing penalties.

## What we see happening at institutions

Most regulated lenders have something going on in the data space. A platform migration, a new BI tool, a small team trying to rationalise reporting. What is less common is a strategy that connects these and ties them to business outcomes.

In our work across the sector, three patterns come up consistently. Based on our experience in the market, roughly half of lenders fall into the first category, around a third into the second and fewer than one in five into the third.

1. **Reactive modernisation.** Something breaks – a regulatory request arrives, an audit finding lands, an MI pack is needed for a board meeting – and the organisation is forced to respond. This produces improvements but rarely a coherent capability. Each fix addresses the immediate problem without reference to a defined target state. The organisation ends up better at the wrong things, or at the right things in the wrong order.

- 2. Technology-led investment.** A platform is selected and the strategy follows. Architecture decisions get made before the business questions have been properly answered: what are we actually trying to achieve and what do we need data to do in three years? The result is often capable infrastructure that the organisation lacks the operating model or governance to use. McKinsey's cross-industry research found that 45 per cent of digital transformations fall short of profit expectations, with only one in ten exceeding them. Their banking-specific work is more telling. As only around 30 per cent of banks surveyed had a data strategy in place before embarking on major platform programmes, with predictably disappointing results. In our experience with UK regulated lenders, the explanation is a bit different: technology investment business cases routinely assume the underlying data is in a far more modern state than it is. When that reality surfaces mid-implementation, timelines extend and the return erodes.
- 3. Strategy in name only.** A document exists, usually produced as part of a transformation programme, but it sits disconnected from how decisions get made. Individual teams choose tools and build processes without reference to any agreed direction. MIT's Centre for Information Systems Research

Found that while 85 per cent of organisations describe data as a strategic asset, only 45 per cent act like it is. That gap is where most data investment gets wasted.

None of this is unique to financial services. But in a regulated environment – where data quality, lineage and auditability carry direct regulatory weight – the consequences are more acute. And as TSB found out, the cost of remediation can dwarf the cost of doing it right the first time.



**85%**

of organisation describe data as an asset and 45% act like it is.

## What a good strategy looks like

These six questions are drawn from our direct experience advising regulated lenders on data strategy – building strategies, assessing why previous ones failed and stress-testing our thinking with data leaders and executives across the sector. What follows is a distillation of what we have seen, that separates strategies driving decisions from those that gather dust.



# Six-question framework

## 1 What are we trying to achieve?

Clear business outcomes, not technology: faster credit decisions, quicker regulatory reporting, earlier risk detection. With this clarity, technology choices follow. Organisations with a data strategy linked to business outcomes are 2.5x more likely to succeed (Gartner).

## 2 Where are we now?

An honest assessment of current capability across architecture, engineering, governance and analytics. Common gaps include manual reconciliation, undocumented pipelines and spreadsheet-driven processes. Poor data quality costs £11m annually on average, excluding regulatory and operational impacts (Gartner).

## 3 Where do we want to be?

A clear target state: consolidated systems, agreed data quality standards, single sources of truth and viable AI use cases. Typically a 3-year horizon with measurable milestones.

## 4 How will we get there?

A phased roadmap, prioritised by value. Start with foundations (core data, governance, reporting), then move to analytics/AI. Smaller, incremental investments are easier to deliver and fund than large, long programmes. Chunking the investment also gives the board regular decision points: continue, adjust, or reprioritise. That is how modernisation sustains momentum and funding over time.

## 5 How will we govern and sustain it?

Ownership assignment, link to budget and embed in business planning. Data spans functions, so it needs a senior owner with cross-functional authority. If nobody owns it and no money follows, it will drift.

## 6 What does success look like?

Measurable outcomes: demonstrating faster reporting, fewer manual processes, increased use of trusted data and live AI use cases. Executive support depends on clear links between data investment and business performance.

## Building your strategy in practice

For most institutions, the starting point is discovery - typically four to six weeks - that assesses the current state honestly, surfaces the priority gaps and produces a sequenced roadmap. That roadmap then becomes the brief for architecture, engineering, governance, analytics and business decisions.

The discovery needs both technical and business input. Strategies built by data teams in isolation tend to reflect what is technically interesting rather than what the business needs. Strategies written by leadership without grounding in the technical reality tend to underestimate dependency, overestimate pace and miss the operating model changes required to sustain what gets built.

In our experience, most lenders already know more about their problems than they give themselves credit for. The data quality issues are understood - someone in finance reconciles them every month. The fragile pipelines are known - the analyst who built them has been meaning to replace them for two years. The spreadsheet-based reporting is known - the risk manager who produces it has raised it three times. What most organisations are short of is a structure to prioritise the response and enough confidence that the direction they choose is the right one.

DBS Bank launched a five-year data and digital strategy in 2014, anchored to business outcomes rather than technology choices, with sustained commitment from the top and a willingness to overhaul its operating model alongside its platforms. By 2018, two years ahead of schedule, it had achieved its goals and was rated among the best banks in the world. The Harvard Business School case study on it identifies those three factors - not budget, not technology selection - as what distinguished it from comparable programmes that fell short.



## The relationship to what follows

The sections that follow - data architecture, engineering, governance and analytics - each address a specific capability area. This document is intended to give institutions a working understanding of each and how they connect.

But capability decisions are only as good as the strategy behind them. Architecture choices made for the wrong objective produce the wrong platform. Governance structures without clear ownership fail in practice. Analytics built on unstable foundations lose credibility quickly and that credibility is hard to recover.

Regulated lenders hold a particular position in the UK financial system. Building societies alone serve around 25 million members, hold £525 billion in assets and accounted for nearly three quarters of all UK mortgage growth in the second half of 2024. The scale of that responsibility warrants data decisions made with rigour.

Most lenders we speak to know they need a data strategy. The harder challenge is building one that remains useful. No strategy lasts indefinitely - it should be treated as a living document with an explicit review cycle, typically annual and clear triggers for reassessment: a core system replacement, a regulatory change, a shift in business model, or a material change in leadership. A strategy that cannot adapt is a strategy that will be quietly abandoned.





Colleagues are organised into tribes, close to the product domains so that they understand the product and business more. This helped to shift the quality and deliverables. Self-service is an addition, giving people access if they want to build on top of what's existing. Having the tribe leaders clarifies the accountability.

Ravneet Shah, CTO, Allica Bank



# Data Architecture

## Introduction

Data architecture is the designed flow of data, from source to consumption. Specifically, it focuses on data modelling, schema definition, governance and determining relationships between systems and domains. This is supplemented by the infrastructure, which outlines the tools, servers and networks that make up this architecture.

Within lending, a good data architecture ensures a structure, scalable and secure platform to combine the various data sources into actionable insights, enabling faster, more accurate credit decisioning. This reduces manual dependencies in the underwriting and assessing process, reducing

overhead costs and improving the customer experience in knowing the outcome sooner.

Key business cases that are dependent on a best-in-practice architecture include:

- Single customer view for personalised engagement
- Full traceability for regulatory and audit compliance
- Operational efficiency through reduced manual reporting
- Clean, accessible data for predictive analytics & AI



## What's currently happening at institutions?

Based on the conversations Woodhurst has had with regulated lenders, the north star for architecture across the sector can be summarised as a combination of:

- Trusted, accessible data that can be reused across the lender's domains
- A single, governed view of core business information
- Less manual handling and fewer tactical extracts
- Better self-service, without losing clearer ownership across domains
- Stronger data quality controls at source
- Architecture that supports flexibility, product change and growth
- Readiness for AI, advanced analytics and automated decisioning
- Faster, safer migration away from legacy dependencies
- Better customer and colleague experiences through smarter reuse of known information

Whilst the starting points and current status of each regulated lender differs, these principles should guide the architecture that is best suited for your organisation.

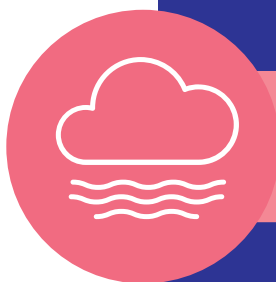
## What does success look like?

Our conversations with regulated lenders have highlighted the following success metrics for a modern data architecture:

- Easy to implement
- Low cost
- Scalable

# Let us explore the common approaches to modern data architecture today

## Architecture Types

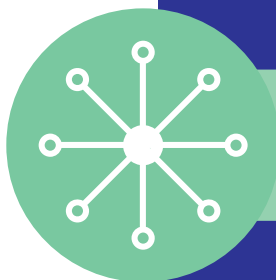


### Data Lake

Centralised Storage of all data types

**What is it?** Flexible storage of all data types (structured and unstructured) designed to be scalable and cost efficient.

**Term popularised in 2010/11**

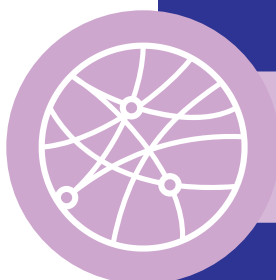


### Data Mesh

Decentralised ownership & governance where data is a product

**What is it?** De-centralised approach where domain (e.g. Finance, Credit) own and manage their data independently.

**Term popularised in 2016/18**



### Data Fabric

Unified data connectivity & management

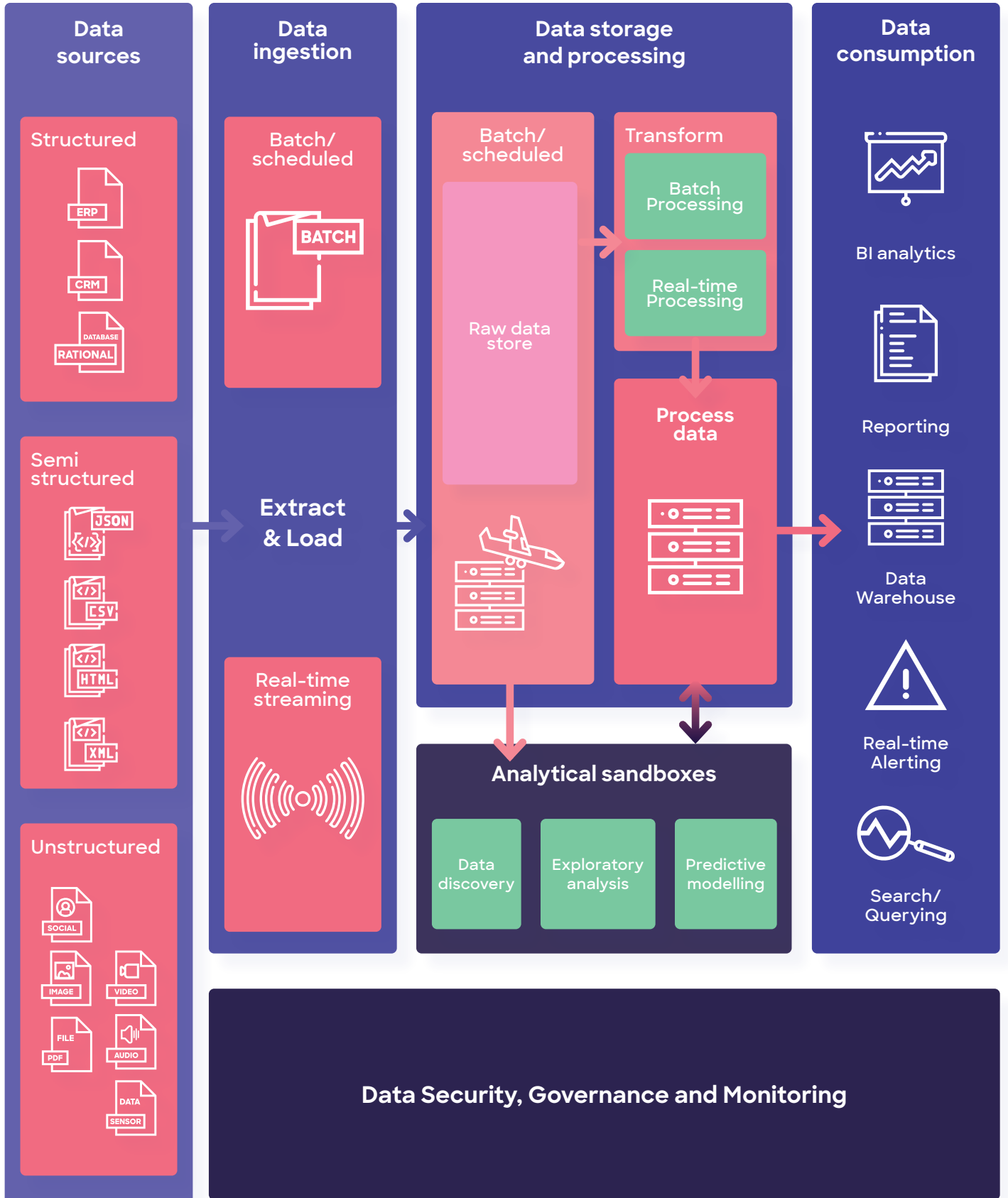
**What is it?** Connects data sources via intelligent layer, leveraging metadata & AI to eliminate silos.

**Term popularised in 2020**

# Lakehouse Considerations

A lakehouse blends low-cost, scalable storage with the structure and performance you expect from a warehouse. In practice, it is the quickest route to getting core data under control: consistent definitions, repeatable pipelines and an auditable path from source to report.

It also aligns well with the reality of smaller data teams. You can standardise ingestion and modelling once, then reuse patterns across mortgages, savings, finance and risk.

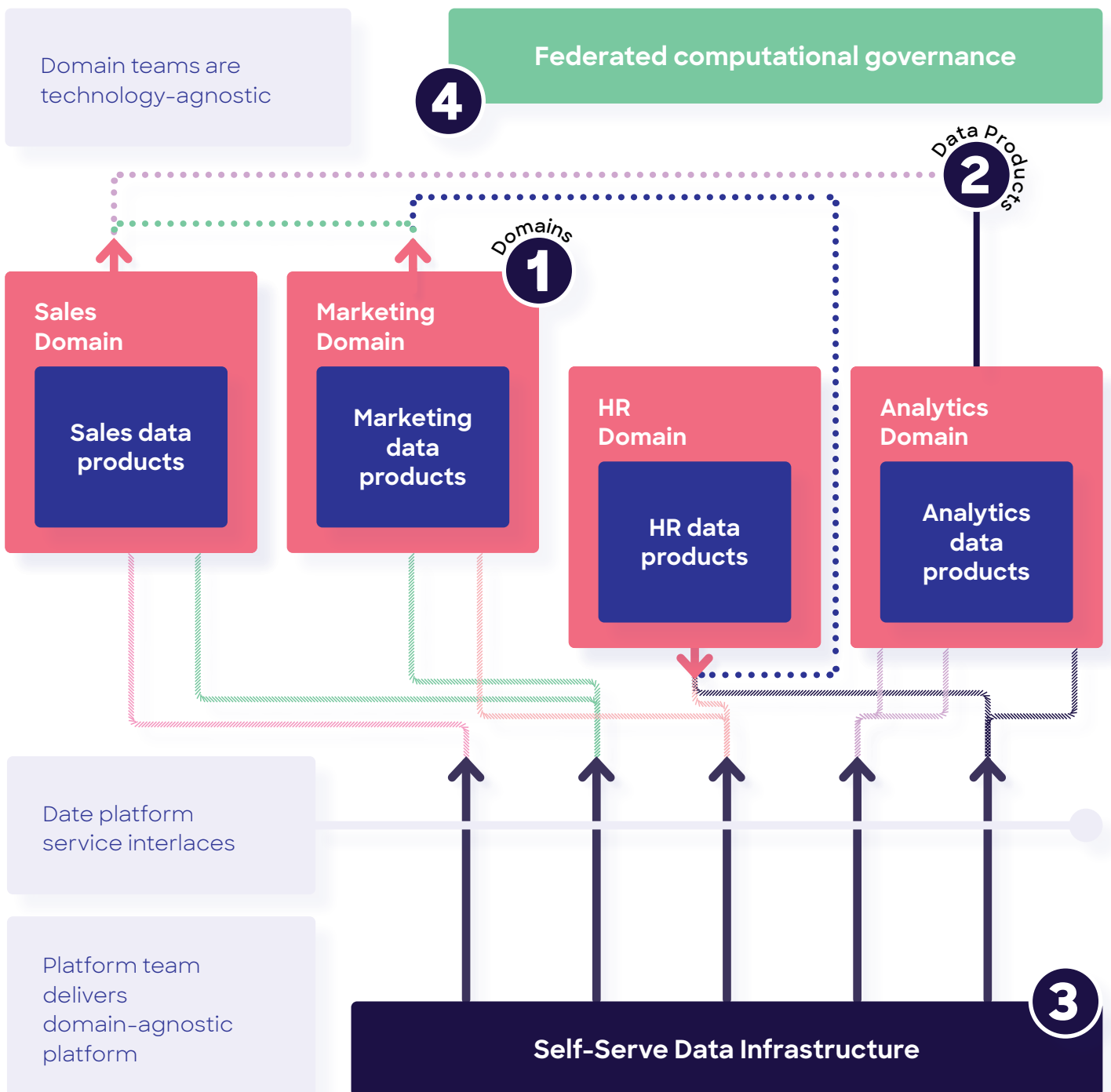


## Data Mesh Considerations

Alternatively, your organisational structure may be best suited to a data mesh. This removes the requirement for a siloed data team, by federating data governance with centralised policies and distributing data ownership (i.e. providing data as a product) within each of your domains, such as mortgages, savings, customer service. This integrates data within operations, with a data SMEs existing throughout the teams and levels of your organisation. That can work well in large organisations with clear domain boundaries and real capacity in the business to take on that accountability.

For most regulated lenders, the risks become fragmented standards, uneven quality and governance that becomes harder, not easier.

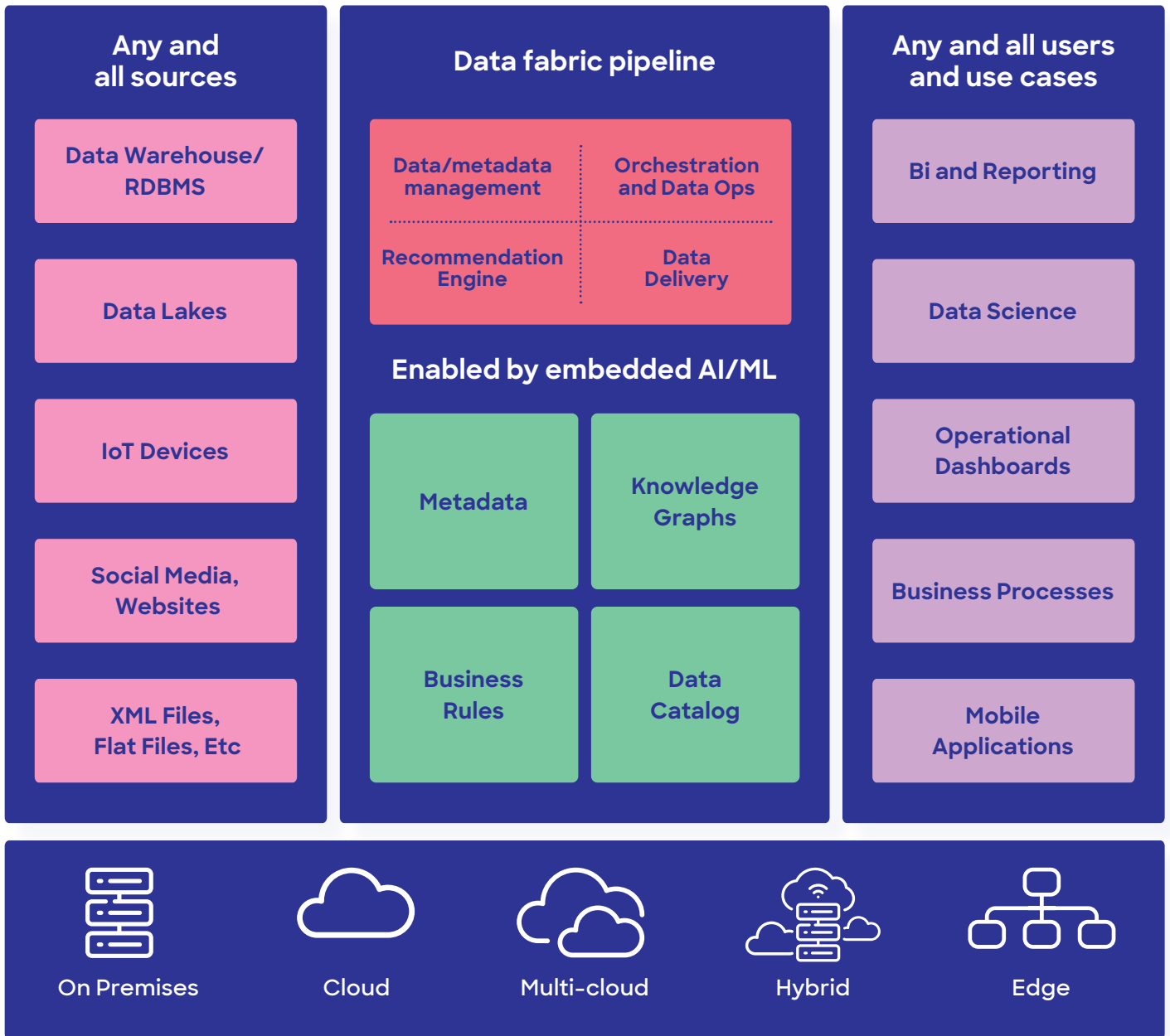
A lender noted that data quality became materially easier to improve once analysts were aligned to business tribes, bringing them closer to products, customers and operational context. This only worked because they had implemented central controls, standards and trusted data products.



## Data Fabric Considerations

Data fabric is best understood as an integration and governance layer rather than a destination in its own right. It can be valuable when you have multiple platforms and a long tail of legacy systems you cannot rationalise

quickly. The issue is that fabric only works when your metadata, controls and operating discipline are already strong. For many lenders, it becomes an expensive overlay on top of unresolved foundation problems.



## So, what should you do...

There is no universal winner, as all approaches have their benefits and can work in conjunction with each other. For example, a lakehouse could be implemented with a semantic layer on top for consumption, with data ownership spread across business domains. A sensible default for most UK regulated lenders would be starting with lakehouse as a central location from which to do reporting. Design it with

strong governance and clear domain models and keep the option open to evolve. Mesh can be a targeted exception where a domain is mature enough to own its data properly. Fabric is a later-stage enhancer once the basics are working. Given the choice of architecture is so specific to the needs and capabilities of your organisation, Woodhurst helps lenders decide on what their ideal model is.

## What Is The North Star?

For lenders, the target-state architecture is one that is trusted, auditable and adaptable. This will ensure you are confident in your numbers, clarity for your regulator and enough flexibility to support future growth without another rebuild.

At its core, the target state is cloud-based and scalable. Data storage, compute and capacity flexes with demand, such that costs are transparent and aligned to actual usage.

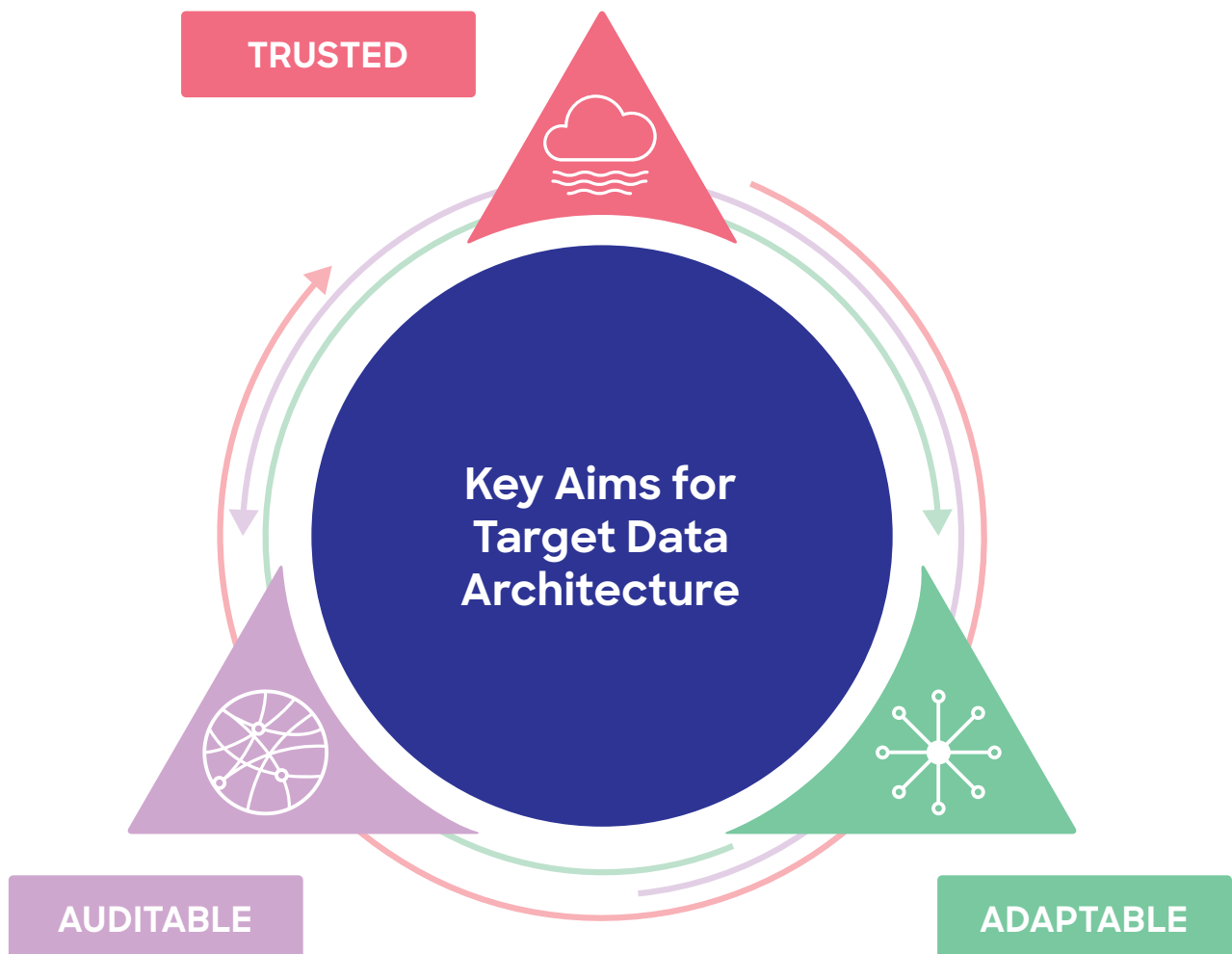
Data is organised in clear, controlled layers. Raw data is ingested once and retained. It is then cleansed, reconciled and standardised before being surfaced as business-ready datasets aligned to domains such as lending, savings, customer and finance. This structure creates a consistent path from source system to regulatory return and is commonly referred to as a medallion architecture which is covered in greater detail in the data engineering section.

Governance is built into the platform rather than managed manually. Access controls,

lineage, data quality rules and retention policies are defined once and enforced systematically. That is what makes compliance with BCBS 239, GDPR and operational resilience expectations sustainable rather than reactive.

Finally, data movement is timely. Change data capture and more frequent pipeline updates reduce reliance on overnight batches. The result is faster credit decisioning, earlier identification of arrears risk and a more current view of the customer.

Reaching this state is phased. Most lenders begin by consolidating onto a single cloud platform and migrating critical regulatory and finance reporting. Curated domain datasets and controlled self-serve access follow. Advanced analytics and AI come later, once quality and governance are stable.



## How to implement a modern data architecture

Today, data modernisation naturally involves upgrading legacy infrastructure to a cloud-native platforms that enable real-time analytics & AI tailored to your organisation's strategic objectives. However, this is no mean feat, given the volumes of data underpinning all lenders and the nuances between platforms. The timelines of implementation of architecture are agnostic to the organisation size, with a large dependency on resourcing. In contrast, the scope of the implementation is specific

to lender, given it needs to consider the current-state, target-state and objectives.

- Discovery & Requirements Gathering
- Design & Proof-of-Concept
- Implementation
- Operationalisation & optimisation

# Phase 0

## Discovery & Requirements Gathering (1-3 months)

Firstly, you want to align data modernisation initiatives to strategic objectives with a clear return on investment. This is ideally documented in your data strategy with alignment throughout leadership. These will serve as guiding principles in decisions to come later. For example, the goal of reducing overheads through automation drives different decisions and outcomes to enabling AI.

Once the target is defined, document the existing data sources and processes. A particular emphasis should be placed on critical reporting, as this will need to be serviced throughout the modernisation journey and immediately after with critical dependency. With these defined boundaries of must-have requirements, you can continue to collect the should-have and could-have requirements. Prioritising requirements will allow you to focus on quick wins alongside the longer-term strategic project.

Leveraging the business requirements, identify the gaps where the current architecture is falling short, technically and operationally.

# Phase 1

## Proof-of-Concept (1-3 months)

Referring to the Phase 0 outputs, your data strategy and your organisation's skillset, determine the appropriate modelling approach and design the technology stack. Tools will need to support your source systems, ETL processes, storage, processing and analytics; along with basic criteria such as scalability, security, usability, to meet your requirements. Platform choices are determined by volume of data, how quickly you need that data and how often you access the data. Keep in mind the existing technology stack and skillsets employed, to leverage existing skills, while also ensuring the objectives of your modernisation are met.

A popular approach that we've followed for a few clients is building a proof-of-concept to prove the business case value upfront, before undertaking the larger transformation. For example, a single customer view that combines data across systems demonstrates how centralised architecture supports daily pricing decisioning.





Data is an emotive subject. Teams have their own data they know well and while everyone supports the enterprise level vision and change, adoption can become harder in terms of the local disruption it can create in the trade-off between organisation alignment and locally embedded processes.

Carole Roberts,  
Director of Technology & Data,  
Leeds Building Society



## Phase 2

### Implementation (6 months)

This phase involves the data engineering element of setting up the architecture you have designed. More details on this in the following 'Data Engineering' section. A phased approach to the migration to the new architecture will allow for gradual adoption and testing in adequate sizing, while the business continues with BAU activities.

## Phase 3

### Operationalisation and Optimisation (Ongoing)

A lender summarised the challenge of adoption well: Teams have their own data silos they know well and while everyone supports change initially when there is an agreed problem, adoption becomes harder when it affects their specific workflows and data. It can be difficult getting people to think at an enterprise level rather than just their own needs, which can be particularly overwhelming when reframing 20+ years of established systems and reporting.

Whilst migrating the data to new platforms, empower the people in your organisation to leverage the benefits through change management, upskilling of new tools and resources and inhibit the culture aligned to the modern data architecture. This could look like democratising the data via self-service analytics, encouraging pricing teams to review pricing models with the newly available data, providing the platform for business users to share ideas as business needs evolve.





We've spent the last four years modernising... building a stable base, so that we can truly harness the power of our data. Everyone wants the next sexy thing, which is AI, but you need to have the right foundations in place to accelerate adoption and that's exactly what we've been working on at Hodge.

Laura Alexander, Head of Data Services, Hodge Bank



# Data Engineering

## Introduction

Data engineering is the discipline that turns architectural design into operational control. If data architecture defines the intended structure of the estate, data engineering is the practical implementation and ongoing management of it.

Data engineering is therefore the mechanism through which data from core mortgage systems, savings platforms, finance ledgers, broker portals and third-party tools is brought into a consistent, governed and reconcilable foundation. It establishes a traceable path from operational activity to board reporting, regulatory submissions and published accounts, improving the efficiency and effectiveness of deriving necessary insights.

Historically, many lenders built their data platforms incrementally. Extracts were generated from core systems, reconciled

in spreadsheets and aligned to finance periodically. Transfers between non-integrated systems were highly manual processes, or had fragile, single function pipelines to reduce colleague overhead. Core metrics were regularly recalculated within each pipeline, with potential for deviations in definition. Whilst this works in the short-term, as the data landscape scales so too does the complexity of the estate - increasing the time it takes to maintain and improve the landscape. Additionally, it limits the trust the business can place in their data - there is no single source of truth.

Modern data engineering introduces structured discipline: repeatable ingestion, layered design, embedded reconciliation logic and preserved history. In a prudentially regulated environment, this discipline underpins defensibility.



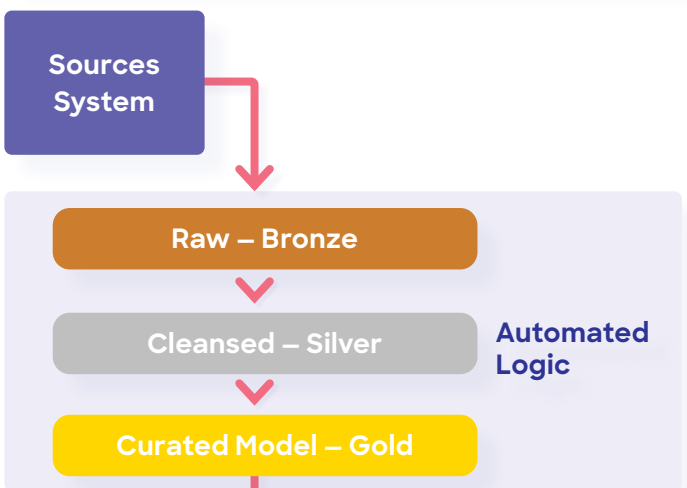
## From Tactical Extracts to Controlled Pipelines

As organisations seek improvements in the utilisation of their data, we see architectures that formalise and consolidate the way that data flows through the firm. This typically involves:

- Automated ingestion from source systems into a sole source, with integrations prioritised over RPA.
- Clear separation of raw, cleansed and curated data. Raw ensures auditability over the data's history, whilst the cleansed and curated layers do the grunt work of most transformation logic to ensure consolidated definitions.

### Controlled Estate (Modern)

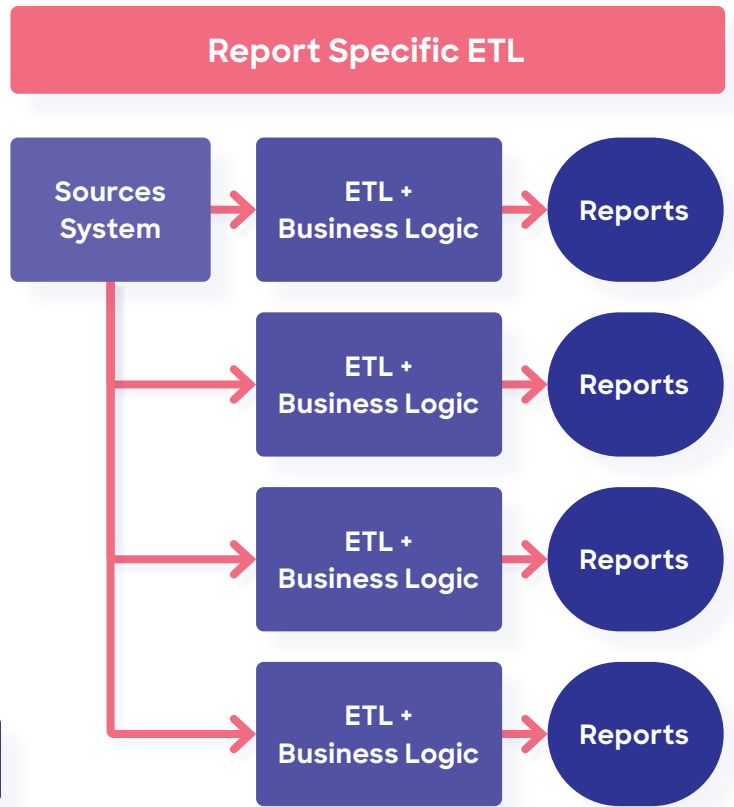
#### Centralised Data Engineering Pipeline



Reports are built on a shared business data model

- Business definitions are centralised
- ETL is centralised, minimising overheads
- Access to data democratised and governance is handled by teams

### Tactical Estate (Legacy)



Each report has its own pipeline and transformation logic

- Business definitions are duplicated
- Duplicated ETL pipeline for similar reports
- Heavy reliance on BI teams for maintenance and new reports

- Structured historical capture to support point-in-time reporting, time series analysis and ease of integration into business intelligence tooling.

These patterns reduce duplicated logic, improve transparency and create a stable foundation for downstream reporting and analytics.

However, adopting these structures raises important questions around how Extract-Transform-Load (ETL) processes are designed, governed and operated.



## Engineering is Pragmatic

Ambition does not always align with institutional reality.

Legacy technologies, prior design decisions and internal capability constraints can limit what is feasible in the short term. At the same time, the cost of building and maintaining a highly sophisticated data estate must be proportionate to the value derived.

### Common tensions include:

#### Real-time aspiration vs. operational necessity

A recurring tension is the push for real-time capability versus actual operational necessity. While more frequent data updates can enhance visibility of pipeline volumes or arrears movements, many core banking systems remain inherently batch oriented. As a result, the marginal benefit of real-time ingestion must be carefully weighed against the associated cost, architectural complexity and increased control risk.

In practice, our experience suggests that the value of real-time insight is often overstated. Market forces are not driving a universal need for real-time capability. Instead, the greatest value is consistently realised through improved data quality, ease of use and enhanced analytical capability on well-governed batch datasets.

#### Modern platforms vs. sustainable operating discipline

Cloud adoption and modern tooling materially improve scalability and flexibility. However, they also raise the baseline for engineering maturity – requiring version-controlled codebases, structured environment separation, CI/CD development processes, automated testing and active monitoring.

This creates a structural tension: while platforms lower the barrier to building advanced data capabilities, they simultaneously increase the organisational burden required to operate them safely and sustainably. Without sufficient engineering capability, institutions risk recreating legacy fragility within a modern architecture – only now at greater scale and speed.

In practice, the challenge is not simply adopting modern tooling but aligning it with the organisation's capacity to sustain

it. Over-investment in platform capability without a commensurate uplift in skills and operating discipline can introduce unnecessary cost, complexity and delivery risk – particularly where existing teams must be upskilled to bridge the gap.

From our experience, leading institutions actively calibrate this balance. Rather than maximising technical sophistication, they right-size their architecture and engineering practices to match their current and near-term capability – prioritising reliability, maintainability and incremental maturity over theoretical best practice.

#### Delivery speed vs. defensibility

Short-term solutions – such as embedding transformation logic within BI tools – may accelerate delivery while the broader data estate is maturing. However, these shortcuts weaken traceability and create technical debt that must eventually be addressed.

For most regulated lenders, the appropriate position is measured progression. Prioritise repeatability, reconciliation and auditability first. Expand sophistication where it delivers clear operational or regulatory benefit.

## The risks of not taking data engineering seriously

Whilst data engineering is pragmatic, firms should not be constrained solely by what is feasible today. The reason lenders undergo modernisation is not to chase modern tooling for its own sake, but to improve competitiveness: to deliver better customer outcomes and reduce operational friction whilst protecting themselves against data-centric new entrants, ever-evolving regulatory expectations and rising scrutiny around model risk, fairness and transparency.

In a data-driven world, investment in a trusted data foundation typically has a short payback period –not because it creates “insight”, but because it reduces waste. It eliminates duplicated reporting effort, decreases manual reconciliation, reduces errors, accelerates change delivery and makes regulatory responses faster and less disruptive. For most regulated lenders, the cost of not modernising increasingly shows up in hidden operational drag: fragmented MI, slow decisioning, inconsistent definitions



across functions and an inability to scale automation without introducing control risk.

Executives therefore face a structural tension. A cautious approach may feel pragmatic in the short term – particularly when current reporting “works”, budgets are tight and engineering capability is scarce. But over time, that caution compounds into strategic vulnerability. Competitors with modern data platforms can launch products faster, adapt underwriting in days rather than quarters, respond to regulatory change with confidence

and use analytics to improve customer outcomes at scale. Meanwhile, traditional institutions remain stuck in an estate where every new requirement becomes a bespoke build and every change introduces new fragility. Finding the right way to modernise data, specifically through data engineering work, should therefore be a top priority.



## How to implement data engineering

Whilst data engineering is the core of the implementation of a new architecture, it is important that data engineers engage in both the requirement capture and proof of concept design. The core to getting data engineering right is ensuring that the engineers are not segmented from the wider business

– the smaller the disparity between what the engineers think the business wants and their actual requirements, the better. Additionally, it is at this time where engineers can bring a pragmatic lens to business requirements, being the balancers between what is doable and what is right for the business.

# Our Checklist

## Step 1

Assess – “Are we set up to succeed?”

---

### *Team and Capability Readiness*

- Engineers are comfortable with version control systems and collaborative workflows.
- Team has knowledge of CI/CD pipelines.
- Engineers understand data modelling principles (facts, dimensions, SCD history)

### *Operating Model*

- Data engineering is not siloed from business stakeholders.
- Engineers participate requirement capture, not just delivery.
- Engineers are enabled to actively challenge and refine business requirements.

# Step 2

Define – “Do we actually know what we’re building?”

---

## *Business Alignment*

- Analytical use cases are clearly defined.
- Reporting expectations are explicit and prioritised.
- Future-state data access models are agreed (self serve, curated reporting, hybrid)

## *Governance and Ownership*

- Access controls are defined (who should see what, at what level)
- Domain ownership is clear (who owns which datasets)

## *Source System Understanding*

- Source system data models and relationships are documented.
- System interactions and dependencies are understood.
- Constraints are known (latency, update frequency, API limits)

## *Target Architecture Design*

- Have we defined what goes into each layer of the architecture?

# Step 3

Implement and Iterate

---

## *Engineering Discipline*

- Pipelines are modular and reusable.
- Workloads use parallelism and concurrency where necessary.
- Cost is monitored and aligned to SLAs.

## *Resilience and Failure Handling*

- There is a clear approach to handling:
  - Source schema changes.
  - Backfills and reprocessing.
  - Pipeline failures and retries.
- “Unhappy paths” evaluated.

## *Migration Strategy*

- Migration is phased, not big bang.
- Legacy reporting continues during transition.
- Data parity between old and new systems is validated.

## *Operational Capability*

- Monitoring and alerts are in place.
- Ownership for pipeline support is defined.
- SLAs and expectation are documented.



“

Governance is key but it's worth highlighting that governance can add friction, so there has to be a right balance struck. There needs to be education on how to protect your core data without hindering progress.

Ravneet Shah, CTO, Allica Bank

”

# Data Governance, Management & People

## Introduction

In the industry today, especially in the regulated lending space, infrastructure is modernising faster than governance considerations, with cloud adoption accelerating and core replacement underway across the sector. In addition, from a regulatory perspective, the shift towards Consumer Duty and increasing regulatory scrutiny is placing greater emphasis on outcomes and accountability, reinforcing the need for robust data governance. This section not only explores establishing data governance but also the lessons learned.

## Governance begins with ownership

At the heart of effective data governance sits accountability.

While everyone in an organisation plays a role in upholding data standards, there must be clarity on formal ownership. While there are lots nodding heads on principles for governance discussions, there lacks clarity on who owns the actions that follow.

A common misconception is that a strong central data team equates to strong accountability. In reality, without business

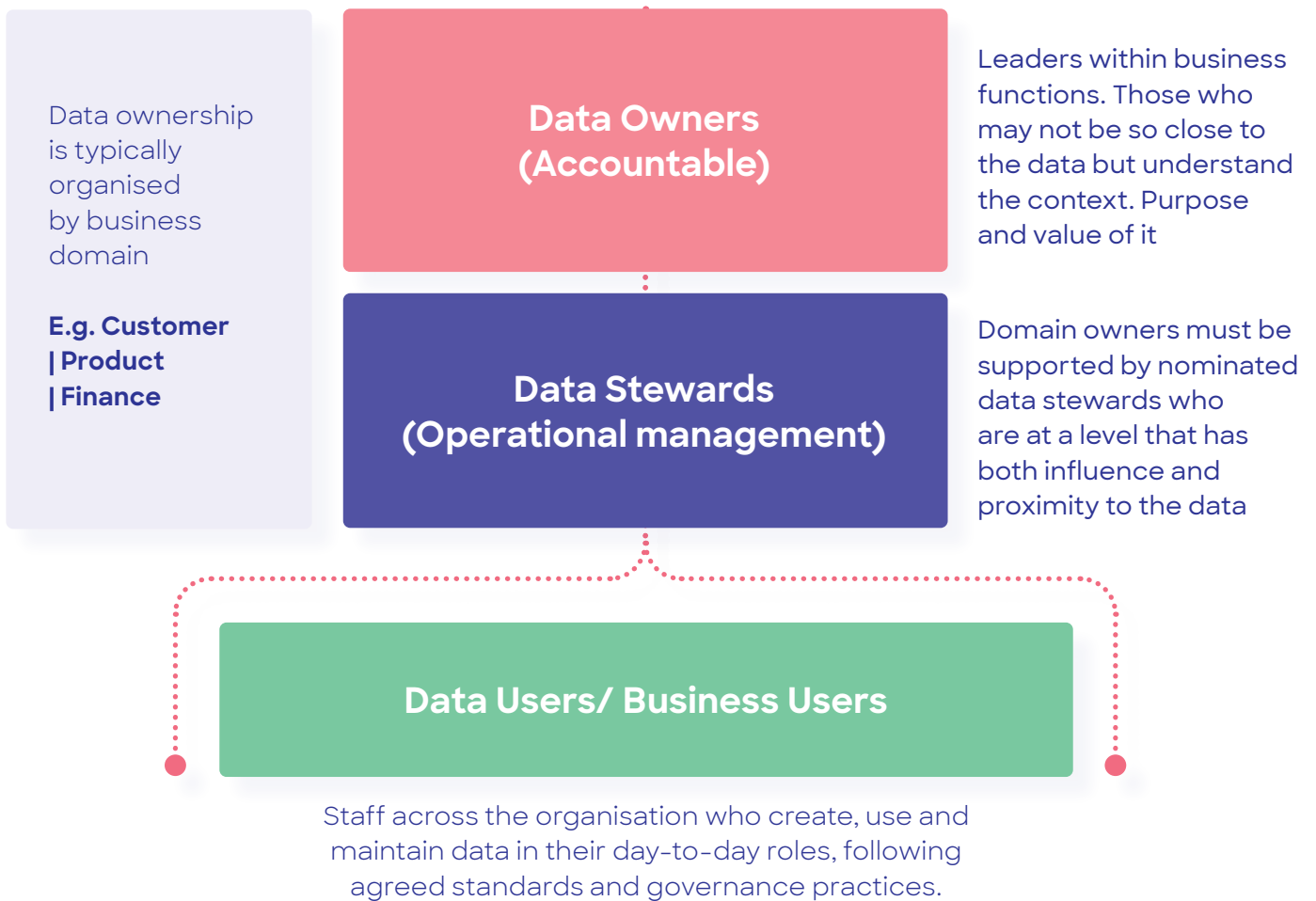


ownership, data risks becoming an operational chore rather than a strategic asset.

Data ownership should sit with leaders within business functions. Those who may not be so close to the data but understand the context, purpose and value of it. Ownership can be structured at a domain level (e.g. customer, finance, product data). These domain owners must be supported by nominated data stewards who are at a level that has both influence and proximity to the data. This overall ownership structure will extend into vendor-led models as vendors will recognise their clients as data owners

**Lenders across the sector agree that ownership cannot simply be assigned through policy, rather reinforced through structure, collaboration and day-to-day operating routines.**

and themselves as data custodians, which is typically a technology function responsible for storage and protection. The same data owner and stewards' formation applies.



A key positive influence in the last decade was the regulatory framework Senior Manager Certification Regime (SMCR) which was introduced to promote accountability and reduce misconduct. This rule is targeting behaviours and decision making and in a growing data driven world, this has had a positive impact on accountability around data governance. The aim of targeting cultural

behaviour and not just documentation is the correct one that tackles data governance at the core of what it should be. This is reflected in practice, where discussions with industry participants consistently highlighted that organisations with clearly aligned executive ownership of data priorities, often through ExCo, have made more tangible progress in their data modernisation journeys.



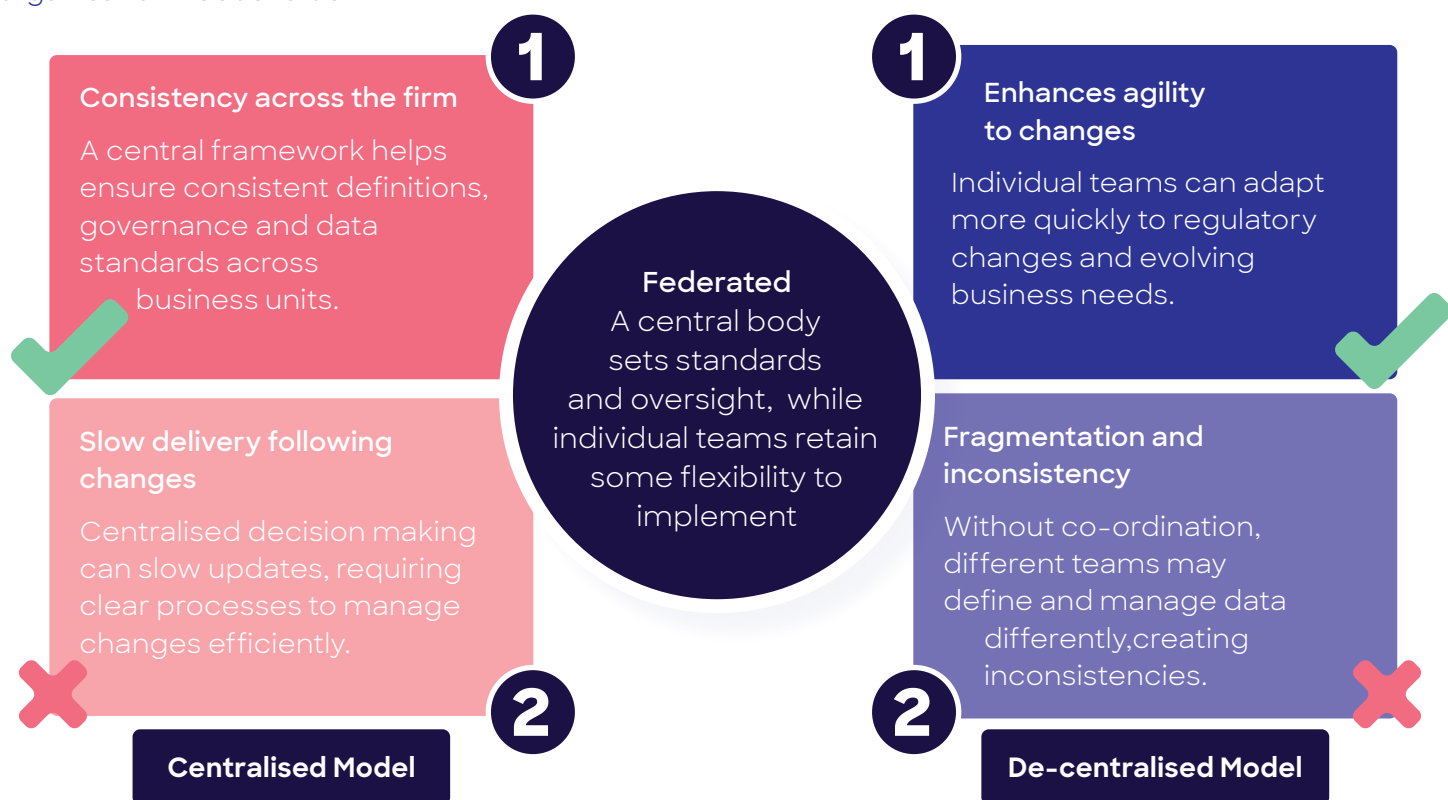
## Maintaining Governance with Changing Operating Model

Operating models need to be reshaped to get value from modernisation. Changes such as Agile delivery, product-aligned teams and domain-based ownership structures are increasingly common. Governance cannot be left behind and must evolve alongside these changes.

Without careful alignment, change can result in conflicting responsibilities or overlapping governance frameworks. The organisation needs to be

clear on where the governance instruction comes from and therefore must determine the appropriate balance between centralised and decentralised governance.

A fully centralised model can ensure consistency across the firm but may slow delivery. A decentralised model enables agility but risks fragmentation and inconsistency. A hybrid approach is often most effective: decentralised domain ownership operating within a centrally defined governance framework that sets minimum standards and controls.



Governance should enable transformation, not constrain it. More governance does not lead to better protection. Strong, clear and agile governance does.

## Governance During Modernisation

Modernisation presents both risk and opportunity. Core replacement and cloud migration in particular expose structural weaknesses in data design.

Modernising legacy structures without redesign risks migrating historical inconsistencies into modern platforms. Schema mismatches, duplicated definitions and fragmented identifiers can undermine integration, reporting and digital enablement.

Lenders should use modernisation as an opportunity to define a common data model,

a consistent way of structuring and defining core data elements such as customer, account, loan, transaction and product.

Before migration, organisations should:

- Agree clear definitions for core data elements.
- Standardise identifiers and status fields.
- Identify and remove duplication and any known business legacy issues.
- Align infrastructure and reporting structures.

Metadata definitions, which is described as data for data, should be treated as architectural prerequisites. Without it, it will create an opaque architecture. In cloud environments, automated metadata harvesting, which is the systematic approach, should be considered essential rather than optional.

Before and during modernisation, poor data governance must be recognised as a silent blocker of success. Too often, organisations treat data as an abstract concern,



**It's logically straight forward to move a core but going through the process can be incredibly complex. You would know the critical data elements, however have you captured what [Person A] is using over there? All of the things you didn't think about, especially when someone may not have documented and if it's a few years old.**

Ewan Mcleod, CTO, Nomo Bank



### Activity

### What is it?



#### Clear Definitions

#### Agree clear definitions for core data elements

Agree consistent definitions for key data elements so systems and teams interpret data in the same way



#### Standardisation

#### Standardise identifiers and status fields

Standardise identifiers, formats and status fields to data can move cleanly between systems



#### Clear Definitions

#### Identify and remove duplications and any legacy data issues

Identify and remove duplicates, outdated fields and poor-quality data migration



#### Infrastructure Alignment

#### Align infrastructure and reporting structures

Align data architecture and reporting structures so the new environment supports future use

overlooking critical dependencies that ultimately delay or derail transformation progress.

While many organisations begin focusing on data during modernisation, there is an opportunity to strengthen outcomes by embedding data clarity and lineage earlier in the process.

While many organisations begin focusing on data during modernisation, there is an opportunity to strengthen outcomes by embedding data clarity and lineage earlier in the process.

## Importance of Transparency and Hygiene

For regulated lenders, expectations around data structure, traceability and quality are driven by the need for defensibility, the ability to evidence how decisions are made, what data supports them and whether outcomes are appropriate.

### Regulatory Expectations and Defensibility

Standards such as BCBS239 and supervisory expectations from the PRA place increasing emphasis on:

- Data aggregation capability
- Accuracy and completeness
- End-to-end traceability

While many institutions can trace individual data flows when required, few maintain a consistent, enterprise-wide view of how data moves across systems and dependencies. This limits their ability to respond confidently to regulatory scrutiny.

### Data Lineage and Change Impact

This lack of visibility becomes particularly challenging during modernisation programmes.

Without clear data lineage, institutions struggle to:

- Assess the downstream impact of system or data changes
- Identify dependencies across platforms
- Manage transformation risk effectively

True data lineage should not be treated as static documentation. It should provide operational visibility, enabling organisations to understand, monitor and evidence how data moves across the estate.

### Data Quality and Customer Outcomes

Data quality expectations have evolved significantly in the regulated lending space. Consumer Duty has continued to shift the emphasis from technical measures such as “accuracy and completeness” to tangible customer outcomes.

As a result:

- Critical Data Elements (CDEs) are more rigorously defined and controlled.
- Data such as vulnerability flags, arrears status and affordability indicators must be consistently accurate.
- Boards are placing greater emphasis on whether the data used to make decisions can be trusted and evidenced.

Customer outcomes and regulatory concerns are portions of market facing issues with data. The broker impact is increasingly more important also. 87% intermediary led lending books according to IMLA's December 2025 Mortgage Market Forecast.



**87%**

of lending books are intermediary led.

### Continuous Data Quality Control

Data quality issues often surface during modernisation, particularly during migration or system integration.

However, data quality should not be treated as a one-off remediation exercise. It must be managed as an ongoing control discipline.

The appropriate level of monitoring depends on the data:

- High-risk, high-velocity data (e.g. transaction or fraud monitoring data) should be monitored in real-time.
- Lower-risk or analytical data (e.g. performance trends) may be monitored periodically.

Data quality will be limited without strong foundations in ownership and structural design, however in controlling it, it is not a one-off remediation exercise. It is an ongoing control discipline that supports, operational resilience, market benefit and customer protection. Data chaos is not just a technology issue, it's a brand risk.

## Governance in a Hybrid and Distributed Estate

Modern lending environments are increasingly hybrid, spanning on-premise systems, cloud platforms and SaaS vendors. This complexity introduces structural and security challenges.

Multiple versions of the same data may emerge across environments. Institutions must define clear roles between data owners (business accountability) and data custodians (IT responsibility for storage and protection). The ambition of a “single version of truth” should represent a joint commitment between business and IT.





From a security perspective, this complexity breeds risk from more entry points, mis-aligned security controls due to capabilities between estates, data in transit and potentially multiple access and identity management systems. Governance should consider consistent protection across estates and must

align with regulatory requirements, including GDPR. Institutions should implement:

- Zero-trust principles
- Encryption
- Data classification frameworks
- Clear retention and disposal policies

A strong starting point is having an overview of the data landscape by use of a data catalogue. Data catalogues should not function merely as searchable libraries. They should operate as control layers, providing visibility, ownership clarity and controlled access to key data assets.

Many breaches come from misconfiguration, not architecture choice. Governance must evolve with the complexity of the landscape.

Principles	Why?
 <p data-bbox="408 1196 571 1229"><b>Zero-trust</b></p>	<p data-bbox="740 1160 1043 1193"><b>Minimise breach risk</b></p> <p data-bbox="740 1207 1370 1263">Minimises breach impact by always enforcing strict identity verification and least-privilege access</p>
 <p data-bbox="408 1397 576 1431"><b>Encryption</b></p>	<p data-bbox="740 1361 1174 1395"><b>Prevent unauthorised access</b></p> <p data-bbox="740 1408 1286 1464">Protect data from unauthorised access by rendering it unreadable without proper keys</p>
 <p data-bbox="408 1599 616 1659"><b>Data Classification</b></p>	<p data-bbox="740 1576 1166 1610"><b>Appropriate protection level</b></p> <p data-bbox="740 1624 1225 1680">Ensure sensitive data receives the right level of protection based on its risk</p>
 <p data-bbox="408 1800 603 1861"><b>Retention and Disposal</b></p>	<p data-bbox="740 1787 1182 1821"><b>Reduce compliance exposure</b></p> <p data-bbox="740 1834 1370 1890">Reduces risk and compliance exposure by securely deleting data when it is no longer needed</p>

## Business Alignment

- Do you have a clear view of your data landscape across on-prem, cloud and SaaS systems?
- Are data ownership roles clearly defined between business (owners) and IT (custodians)?
- Is there alignment between business and IT on maintaining a single version of truth?
- Have you identified where multiple versions of data exist and addressed them?
- Do you use a data catalogue that provides visibility, ownership and controlled access (not just search)?

## Security, Control and Compliance

- Are security controls applied consistently across all environments?
- Have you mitigated risks from multiple entry points and system integrations?
- Is data protected both in transit and at rest (encryption)?
- Do you have a clear identity and access management approach across platforms?
- Are data classification, retention and disposal policies defined and enforced?
- Are your practices aligned with regulatory requirements (e.g. GDPR)?
- Do you use a data catalogue that provides visibility, ownership and controlled access (not just search)?

## Practical Effectiveness

- Are governance issues (e.g. breaches or errors) more often linked to misconfiguration, and are these actively managed?
- Does your governance approach evolve alongside increasing technology complexity?
- Do you have consistent control and oversight of data, regardless of where it resides?

## Data Culture and People

Institutions have traditionally treated data quality as a back-office responsibility. This approach fails. Those who create and use data understand it best and no single function can manage it effectively in isolation.

Data must be recognised as a strategic asset and competitive advantage. Governance is therefore a business imperative, not a support function. Training alone is insufficient. A strong data culture must be embedded in policy, standards and operating models to drive initiative-taking rather than reactive behaviour.

Too often, data stewards are brought in late and asked to govern retrospectively, leading to disengagement. Governance must be done with the business, not to it. Stewards should be involved early, particularly in designing standards and controls, to ensure ownership and buy-in.

The leadership function of an institution need to reinforce the culture to obtain staff buy-in. The example approaches that can be used are establishing KPIs, funding steward capacity and providing authority and creating more dialogue around data quality in forums. The incentive of the leadership team is the implementation of this, will result in overall increased data quality, reduced risk, trusted product management and increased customer trust.





Modernisation has been about creating a solid foundation. With that now in place, we've shifted from maintaining core operations to actively harnessing data as a driver of innovation and value.

Laura Alexander, Head of Data Services, Hodge Bank



# Data Analytics & AI

## Introduction

Data analytics is often described as the top of the data stack. In reality, it is the moment of truth where many data modernisations fail. For most business users, analytics is all they see. It is the dashboard opened each morning, the MI pack presented at ExCo, the figures submitted to the regulator. The sophistication of the underlying platform is invisible and, to them, often irrelevant. If insights are unclear or metrics disputed, the credibility of the entire data programme is called into question.

For many mid-sized financial institutions, the ambition for analytics remains too modest. Considerable effort is often expended simply trying to produce an accurate, reconciled figure. While this should be the priority, too few organisations are asking themselves what deeper insight might be possible. This section will explore the journey of data analytics and how to take it further.

## Data Analytics Today

Analytics has come a long way in the last decade and mid-sized organisations are gradually catching up to the modern way of doing things, usually having developed their analytics capabilities in an ad-hoc rather than joined-up way. While the general concept of a data warehouse (i.e. centralising your data for easy reporting) is not new, these organisations often evolved a patchwork of tactical reporting solutions including isolated dashboards and spreadsheets, direct extracts from source systems and heavy transformation logic in BI tools.





Today, the industry standard for analytics is moving towards a clear separation of duties: a centralised data platform (such as Fabric, Databricks or Snowflake) which ingests data, holds the bulk of transformations and surfaces curated datasets in a clean data model to BI tools. This modern architecture was explored



in earlier sections, but its benefits become clear at the point of analysing clean, well-structured data. Rather than embedding business logic within dashboards, organisations can define core metrics once and reuse them consistently across reporting and modelling. In doing so, they shift from producing isolated

reports to building governed, reusable data assets that scale with the business.

There are several modern, enterprise-grade BI tools available in the market, four of which are summarised here:

Tool	Presence in UK FI Mid-Tier	Typical Fit	Summary
 Power BI	High	Organisations already invested in Microsoft ecosystem	Dominant in UK FS due to cost efficiency, Microsoft integration and strong skills availability. Success depends on disciplined data modelling.
 Tableau from Salesforce	Medium	Organisations with mature analytics teams	Strong exploratory and visual analytics capabilities. Typically, higher licensing costs.
 Qlik	Low	Organisations with established Qlik estates and internal expertise	In-memory associative engine that enables dynamic exploration across linked datasets. More prevalent in legacy estates than in new cloud-first builds.
 Looker	Low	Organisations using Google Cloud Platform	Built around a central semantic modelling layer that enforces metric consistency. Less common in UK mid-sized FS but strong where GCP is the cloud provider.

In practice, tool selection is typically shaped by ecosystem alignment, cost and internal skills profile. The strength of the underlying data model, governance framework and BI team capability will usually have a greater impact on analytics outcomes than the choice of visualisation tool itself.

## Taking Analytics Further

As lenders strengthen their reporting foundations, the opportunity expands beyond standard dashboards and KPIs into more advanced analytical methods and AI-enabled capabilities.

Advanced analytics and data science techniques are designed to uncover patterns, similarity and drivers within data that are not immediately visible through standard reporting,

ultimately giving them the power to answer much more interesting questions. Rather than simply displaying performance metrics in a structured way (telling us what happened), they use statistical modelling and machine learning to perform exercises like group similar entities, quantify relationships, infer characteristics and test how different variables interact.

For a mid-sized lender, practical use cases might include:

- Behavioural segmentation: grouping borrowers by repayment patterns to tailor collections strategies and intervene earlier with higher-risk cohorts.
- Geographic inference modelling: training models on higher-level demographic or behavioural data and applying.



them at a lower geographic level to support more precise targeting of new customers or a new branch location.

- Operational volume forecasting: predicting application inflows by product, channel or broker to align underwriting capacity, reduce SLA breaches and avoid burnout or overtime costs during peak periods.
- Lookalike or similarity analysis: identifying new areas or populations structurally similar to current high-performing customer portfolios to focus distribution and marketing spend more effectively.
- Portfolio drift detection: monitoring how newly approved cases differ from historic lending profiles, such as shifts in incomes or LTV, to flag emerging risk trends before arrears metrics deteriorate.
- Driver analysis: quantifying which variables most influence outcomes such as early redemption, arrears emergence or broker conversion rates, enabling targeted pricing adjustments, policy refinement, or relationship management.

Alongside these techniques, artificial intelligence is becoming increasingly embedded within modern analytics platforms and teams. AI assistants can generate code, suggest features, identify anomalies and summarise findings, lowering the barrier to experimentation and speeding up delivery. Going further, AI agents are becoming able to perform end-to-end data science projects with a high degree of independence. For example, Databricks' Data Science Agent can plan activities, build models, write and debug code and visualise and interpret results, all in response to natural language prompts. This functionality, like any AI agent, relies on a combination of modern tooling and clean, well-structured data. The insights it generates are only as accurate as the underlying data.



## The Prompt

Using this dataset `arrears_data`, identify the key drivers of 90-day arrears over the next 3 months and build a predictive model to flag which currently performing customers in `current_customers` are most at risk. Explain the drivers and quantify the uplift in risk for high-risk segments.

## Planning Activities

To address your request, I'll approach it in three structured stages:

- **Identify Key Drivers:** I'll analyze `arrears_data` to determine which features most strongly correlate with future 90-day arrears (`arrears_90d_next_3m`). Statistical summaries

and feature importance from a predictive model (such as logistic regression or tree-based model) will be used.

- **Build Predictive Model:** Using the key features, I'll train a model on historical arrears data to predict risk. Then, apply this model to `current_customers` to flag at-risk customers.
- **Explain Drivers & Quantify Uplift:** I'll summarize which features are most predictive, provide quantitative estimates of risk uplift for high-risk segments, and highlight actionable insights.

## Writing Code, Debugging, and Producing Outputs

```
%python
SELECT
  arrears_90d_next_3m,
  AVG(missed_payments_last_6m) AS avg_missed_payments,
  AVG(max_days_past_due_last_12m) AS
# avg_1
AVG(:
# avg_2
%python
AVG(:
  from pyspark.ml.feature import VectorAssembler
  from pyspark.ml.classification import LogisticRegression
# Load arrears_data for training
train = spark.sql("SELECT * FROM workspace.bank_arrears.arrears_data")
COUNT
FROM wk
GROUP 1
# Select key drivers
features = ["missed_payments_last_6m", "max_days_past_due_last_12m", "loan_to_income_ratio",
"credit_score", "months_since_last_missed_payment", "interest_rate"]
assembler = VectorAssembler(inputCols=features, outputCol="features")
train = assembler.transform(train)
# Fit logistic regression
lr = Log Here's a quantitative comparison between customers who did and did not enter 90-day arrears in the next 3 months in
model =
# Apply
curr = :
curr = :
```

Segment	Avg Missed Payments	Max Days Past Due	Loan-to-Income Ratio	Credit Score	Months Since Last Missed	Interest Rate	Count
No Arrears (0)	0.20	7	0.43	679	81	0.06	1793
Entered Arrears (1)	1.38	54	0.52	664	2.2	0.07	207

## Presenting and interpreting results

Based on the data in `arrears_data`, the top ranked drivers of 90-day arrears, along with their quantitative impact, are:

- **Missed Payments in Last 6 Months:** Customers who fall into arrears typically have an average of 1.38 missed payments versus just 0.2 for those who remain current—a more than 7x increase.

Here are the top 10 highest-risk current customers in `current_customers`, based on key arrears drivers:

Customer ID	Missed Payments (6m)	Loan-to-Income Ratio	Credit Score	Months Since Last Missed	Interest Rate
2952	3	0.811	738	1	0.0502
2493	3	0.682	643	1	0.0724
4939	3	0.505	604	5	0.0845
4046	3	0.502	590	1	0.0902
3779	3	0.478	749	1	0.05
2574	3	0.464	769	0	0.05
4441	3	0.441	574	2	0.0615
4255	3	0.388	693	5	0.0526
2326	3	0.367	722	2	0.0552
4377	3	0.323	556	2	0.0817

- **Max Days Past Due (Last 12 Months):** Those entering arrears averaged 54 days past due, compared to only 7 days for others. This

highlights the significance of recent severe delinquency.

- **Loan-to-Income Ratio:** High-risk segments have a ratio of 0.52 versus 0.43 in lower-risk groups, indicating greater financial strain.
- **Credit Score:** Arrears customers have a lower average score (664 versus 679), showing creditworthiness matters but is less predictive than recent behavior.
- **Months Since Last Missed Payment:** Time since last missed payment drops sharply to 2.2 months in arrears cases versus 81 months for performing customers—recent missed payments are a strong indicator of risk.

In summary, missed payment history and recent severe delinquency are the strongest predictors, with affected customers showing >2x to >7x uplift in risk compared to the population average. These insights can guide targeted interventions and refined risk strategies.



## Should you enable self-serve?

As analytics capability matures, one of the most persistent debates is how much to delegate it to the business. Should reporting sit firmly within a central BI function, or should business users be empowered to build and manage their own outputs? Many organisations aspire to self-serve to reduce bottlenecks, shorten turnaround times, increase data literacy and allow domain experts to explore their own questions without waiting in a central queue.

In practice, however, self-serve can quickly lead to “cottage industries” - a fragmented reporting environment without proper standards and governance. KPI definitions may diverge, logic can become embedded within individual dashboards and multiple versions of the truth can emerge.

For most mid-sized lenders, the most effective approach is a structured hybrid model. Core data models, KPI definitions and regulatory reporting remain centrally owned to ensure consistency and auditability. On top of this foundation, business users are granted controlled access to certified datasets for exploratory analysis and local reporting. Some organisations adopt an “open by default” model for curated data, reducing friction and encouraging transparency, but this works only where data classification is clear and sensitive information is appropriately protected.

Typical issues with self-serve, alongside effective mitigations, are listed below:

Common Issue	Mitigation
Inconsistent KPI definitions	Centralised semantic layer with agreed metric definitions
Logic embedded in dashboards	Push transformation logic upstream into curated data models
Limited data literacy	Targeted training and certified datasets
Unclear data ownership	Defined dataset owners and governance framework

## How to uplift your Analytics

Lenders can start with identifying the 10 reports that genuinely run the business and rebuilding them in a modern visualisation tool to be clean, accurate, visually appealing and easily accessible. This will build momentum and buy in stakeholders across the business to the new vision. Once this solid foundation is established, further use cases can be explored, initially focusing on high-value, low effort endeavours to maintain momentum.

Following a solid foundation, more complex use cases become a possibility and firms can expand their ambition in a controlled way.

This may be through expanding self-serve, allowing business areas to define and drive their own initiatives to answer more specific questions. Alternatively, complex use cases may justify the addition of more specialist capability, whether through hiring dedicated data science resource or using external partners. The progression should be deliberate: building depth in the fundamentals and expanding responsibly when the organisation is ready to convert insight into action.





# Looking Ahead

## Your Data Modernisation Journey

Here we set out a roadmap for levelling up your data capability. To summarise this approach, an organisation first needs to ensure their attitude and philosophy towards data is conducive to success: setting an explicit commitment to data (via a strategy). This commitment will provide a solid foundation for the modernisation and manifest in the successful delivery of the strategy and associated activities.

After this has been set, a data maturity assessment provides a clear view of the current status, outlining where to direct efforts to ensure they deliver the most impact and provide a transparent view of strengths and weaknesses. A simple self-assessment across the core dimensions of strategy, architecture, engineering, governance and analytics can be used as a starting point to provide a baseline maturity level:

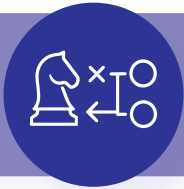
**Leverage our maturity model which can be found in the QR code for your initial assessment.**



This information can then be leveraged to set the direction and define objectives. Each organisation will be at a different point in its journey, but there are common factors that each can apply to ensure success.

## Principles

## Why?



### Strategy

#### Objective linked strategy

Defining a strategy that connects data objectives to business outcomes.



### Collaboration

#### Technical and business input

Ensuring both technical and business input is considered at all points of this journey.



### Timeframe

#### Milestone-led timelines

Setting a realistic time and vision to work towards with clear milestones that can be reported on.



### Responsibility

#### Clear ownership

Defining clear ownership and accountability.



### Continuous Improvement

#### Iterative approach

Creating a loop of continuous improvement - ensuring all members across the organisation can surface issues that can be logged, triaged and prioritised.

## What comes next?

One principle remains constant across all technological advancement: the effectiveness of any innovation is determined by the quality, accessibility and governance of the data that powers it. For organisations that have complete confidence in their data capability, Woodhurst recognises three areas to explore additional value, based on our work with regulated lenders across the market: AI & Automation, Technology Platforms & Ecosystems, and Emerging Financial Infrastructure.

### AI & Automation

Artificial intelligence is rapidly moving from experimentation to operational deployment across financial services, with the House of Commons' AI in Financial Services report noting that 75% of UK financial services firms are using AI. The question is no longer whether AI will play a role, but how to deploy it responsibly within a regulated environment.



**75%**

of UK financial services firms are using AI.

Practical use cases are already emerging across the sector (as explored in our report *Approaching AI within Financial Services - Woodhurst Consulting*):

- Credit decisioning support, improving speed and consistency of underwriting.
- Customer vulnerability detection, identifying early indicators of financial stress.
- Operational automation, reducing manual effort in servicing, reporting and regulatory preparation.
- Customer service optimisation, including conversational and voice AI solutions supporting call centre and IVR functions.

Institutions that attempt to deploy AI on unstable data foundations frequently encounter issues around trust, regulatory defensibility and operational adoption. Several participants suggested that AI has not changed the fundamentals of good data management, but it has made weak foundations far more visible and commercially risky.

The most effective strategy is typically controlled expansion: identifying a small number of high-value use cases and implementing them on top of trusted datasets and governed pipelines.

Commencing each initiative with a small checklist will allow an organisation to quickly understand if they have the required pre-requisites to extract the most value for an AI use case:

- Clearly defined and governed critical data elements.
- Reliable, monitored data pipelines with traceability from source to consumption.
- Established data ownership and stewardship across business domains.
- Model governance frameworks supporting explainability and auditability.
- Controlled environments for testing and deploying use cases.

### **Technology Platforms and Ecosystems**

Ensuring a clear understanding of architecture, governance and operating model before selecting technology tools accelerates progress rather than introducing additional fragmentation. Technology vendors will always play a role in a modernisation journey, however, vendor adoption should follow a strategy, rather than define it.

### **Emerging Financial Infrastructure**

Technologies such as distributed ledger infrastructure and decentralised finance are still developing within the regulated lending ecosystem. While adoption across UK lenders remains limited today, certain applications, such as secure data sharing, digital identity frameworks or settlement processes, may become increasingly relevant in the coming years. For most institutions, these developments remain an area to monitor rather than a near-term transformation priority.

## **The Role of Strong Data Foundations**

Importantly, none of these opportunities exist in isolation. Artificial intelligence, automation, vendor platforms and emerging technologies all depend on the same underlying capabilities:

- Trusted and accessible data.
- Clear ownership and governance.
- Robust and repeatable engineering pipelines.
- Architectures designed for scale and flexibility.

Institutions that establish these foundations place themselves in a position to adopt modern technologies with confidence and speed, delivering tangible benefits across a defined payback period. Those that do not often find innovation constrained by operational complexity and regulatory risk.

For regulated lenders navigating increasing regulatory scrutiny, evolving customer expectations and intensifying competition, the challenge is not recognising the importance of data but embedding it as a strategic capability across the organisation. In practice, this means moving beyond technology programmes and ensuring data is reflected in decision-making, funding allocation and performance measurement. Leading institutions achieve this by aligning executive ownership, linking data initiatives directly to business outcomes and embedding data accountability within operational roles rather than central teams alone.

Organisations that treat it as such will be best positioned to adapt, innovate and grow in the years ahead.

For institutions looking to take the next step, the starting point is often a structured, evidence-based view of their current position and a clear, pragmatic roadmap forward. Woodhurst works with regulated lenders to assess, prioritise and deliver data modernisation initiatives that are grounded in real-world constraints and focused on measurable outcomes.

If you would like to explore how these principles apply within your organisation, we welcome a conversation.



# Authors

Head of Data  
**Sam Spacey**



Experienced data leader with a strong track record of delivering large-scale data modernisation and reporting programmes across complex organisations. Sam specialises in aligning data strategy with business outcomes and has led multiple enterprise-wide initiatives across the sector.

[Sam.spacey@woodhurst.com](mailto:Sam.spacey@woodhurst.com)

**Charles Agunbiade**



Data professional with experience in embedding robust data governance practices. Charles focuses on ensuring data is accurately controlled, managed, and migrated across systems during modernisation, supporting consistency, quality, and trust in enterprise data.

[Charles.agunbiade@woodhurst.com](mailto:Charles.agunbiade@woodhurst.com)

**Gabriella Dixon**



Experienced in designing data architectures, reporting models, and maturity assessments to enhance data governance and enable strategic decision-making across the enterprise.

[Gabriella.dixon@woodhurst.com](mailto:Gabriella.dixon@woodhurst.com)

**Lewis Farrell**



Data Engineer & Platform Architect specialising in end-to-end data modernisation. Deep expertise in cloud-native platform design, DevOps, and enterprise data strategy - implementing platforms that are efficient and built to last.

[Lewis.farrell@woodhurst.com](mailto:Lewis.farrell@woodhurst.com)

**Jake Connor**



Certified Databricks Data Analyst specialising in creating meaningful insights from complexity. Experience with Power BI, SQL, Python, Databricks.

[Jake.connor@woodhurst.com](mailto:Jake.connor@woodhurst.com)

**Charlie Patterson**



Experience in validation exercises and testing, specialising in process automation and AI implementation to support the delivery of an effective data strategy.

[Charlie.patterson@woodhurst.com](mailto:Charlie.patterson@woodhurst.com)





```
function add(a,b){return a+b;}
function sub(a,b){return a-b;}
function mul(a,b){return a*b;}
function div(a,b){return a/b;}
function mod(a,b){return a%b;}
function pow(a,b){return a**b;}
function sqrt(a){return Math.sqrt(a);}
function log(a){return Math.log(a);}
function exp(a){return Math.exp(a);}
function sin(a){return Math.sin(a);}
function cos(a){return Math.cos(a);}
function tan(a){return Math.tan(a);}
function atan(a){return Math.atan(a);}
function atan2(y,x){return Math.atan2(y,x);}
function asin(a){return Math.asin(a);}
function acos(a){return Math.acos(a);}
function atanh(a){return Math.atanh(a);}
function acosh(a){return Math.acosh(a);}
function asinh(a){return Math.asinh(a);}
function acosh(a){return Math.acosh(a);}
function sinh(a){return Math.sinh(a);}
function cosh(a){return Math.cosh(a);}
function tanh(a){return Math.tanh(a);}
function atanh(a){return Math.atanh(a);}
function sinh(a){return Math.sinh(a);}
function cosh(a){return Math.cosh(a);}
function tanh(a){return Math.tanh(a);}
function atanh(a){return Math.atanh(a)}
function sinh(a){return Math.sinh(a)}
function cosh(a){return Math.cosh(a)}
function tanh(a){return Math.tanh(a)}
function atanh(a){return Math.atanh(a)}
```

```
function add(a,b){return a+b;}
function sub(a,b){return a-b;}
function mul(a,b){return a*b;}
function div(a,b){return a/b;}
function mod(a,b){return a%b;}
function pow(a,b){return a**b;}
function sqrt(a){return Math.sqrt(a);}
function log(a){return Math.log(a);}
function exp(a){return Math.exp(a);}
function sin(a){return Math.sin(a);}
function cos(a){return Math.cos(a);}
function tan(a){return Math.tan(a);}
function atan(a){return Math.atan(a);}
function atan2(y,x){return Math.atan2(y,x);}
function asin(a){return Math.asin(a);}
function acos(a){return Math.acos(a);}
function atanh(a){return Math.atanh(a);}
function acosh(a){return Math.acosh(a);}
function asinh(a){return Math.asinh(a)}
function acosh(a){return Math.acosh(a)}
function sinh(a){return Math.sinh(a)}
function cosh(a){return Math.cosh(a)}
function tanh(a){return Math.tanh(a)}
function atanh(a){return Math.atanh(a)}
```

Copyright The Closure Library Authors.  
SPDX-License-Identifier: Apache-2.0

```
var strictFunctionAdder = function(b) { return function(a) { return a + b; }; };
var m = function() { return Object.defineProperty(Object.prototype, 'add', { value: function(a, b) { return a + b; } }); };
function isObject(obj) { return typeof obj === 'object' || typeof obj === 'function'; }
var c = function(a, b) { return a + b; };
function isNumber(a) { return typeof a === 'number'; }
function isString(a) { return typeof a === 'string'; }
function isArray(a) { return Array.isArray(a); }
function isFunction(a) { return typeof a === 'function'; }
```

Copyright The Closure Library Authors.  
SPDX-License-Identifier: Apache-2.0

```
var strictFunctionAdder = function(b) { return function(a) { return a + b; }; };
var m = function() { return Object.defineProperty(Object.prototype, 'add', { value: function(a, b) { return a + b; } }); };
function isObject(obj) { return typeof obj === 'object' || typeof obj === 'function'; }
var c = function(a, b) { return a + b; };
function isNumber(a) { return typeof a === 'number'; }
function isString(a) { return typeof a === 'string'; }
function isArray(a) { return Array.isArray(a); }
function isFunction(a) { return typeof a === 'function'; }
```

```
function add(a,b){return a+b;}
function sub(a,b){return a-b;}
function mul(a,b){return a*b;}
function div(a,b){return a/b;}
function mod(a,b){return a%b;}
function pow(a,b){return a**b;}
function sqrt(a){return Math.sqrt(a);}
function log(a){return Math.log(a);}
function exp(a){return Math.exp(a);}
function sin(a){return Math.sin(a);}
function cos(a){return Math.cos(a);}
function tan(a){return Math.tan(a);}
function atan(a){return Math.atan(a);}
function atan2(y,x){return Math.atan2(y,x);}
function asin(a){return Math.asin(a);}
function acos(a){return Math.acos(a);}
function atanh(a){return Math.atanh(a);}
function acosh(a){return Math.acosh(a);}
function asinh(a){return Math.asinh(a)}
function acosh(a){return Math.acosh(a)}
function sinh(a){return Math.sinh(a)}
function cosh(a){return Math.cosh(a)}
function tanh(a){return Math.tanh(a)}
function atanh(a){return Math.atanh(a)}
```

Woodhurst



# Get in Touch

We believe that everyone  
should have access to  
better financial solutions



Connect with us



Ben Nadel, Co-founder  
[ben.nadel@woodhurst.com](mailto:ben.nadel@woodhurst.com)

(+44) 7498 370 756