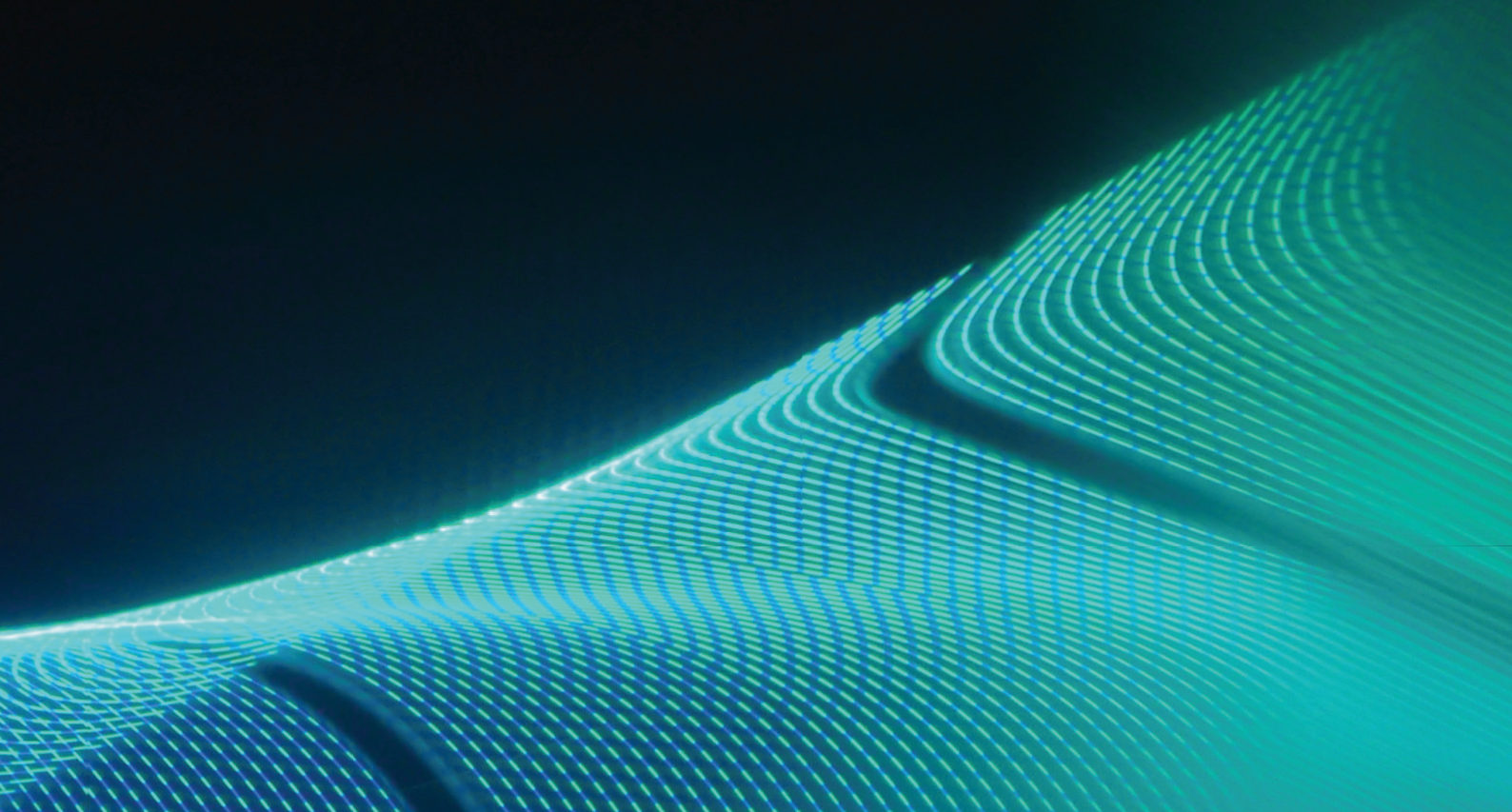
Mastek
Trust. Value. Velocity

Lead with AI

BUILDING A CONNECTED NATIONAL HEALTH DATA CAPABILITY



Executive Summary

The national health data estate needs to be managed as a connected system

The health data and analytics estate is undergoing significant transformation across national, regional and local services. The strategic challenge is ensuring these changes operate as a connected system, strengthening resilience, reducing complexity for the teams that manage and use data, and improving the flow of trusted information to clinicians, operational leaders and patients.

The underlying challenge is structural. Fragmented infrastructure, duplicated processes, inconsistent data collection, resource-intensive reporting and ageing technology continue to create cost, complexity and operational burden across the health system. At the same time, expectations are rising for measurable improvement in clinical outcomes, patient safety, service performance and staff productivity. The longer-term direction is towards more federated access to data, with greater local control and less dependence on central collection, but the current estate remains some distance from that model and must continue to operate safely in the meantime.

This creates a real investment constraint. Limited resources have to sustain critical services, reduce the cost and risk of the legacy estate, and support changes that deliver measurable improvement in care and operational performance. Rationalisation can reduce technical debt and release capacity, but the benefits may take time to appear in frontline outcomes. Investment focused only on new capability can leave an increasingly costly legacy base untouched. Sequencing therefore matters. Simplification needs to create headroom for improvement, while new capability should reduce future dependence on legacy technology.

That challenge cuts across the full data lifecycle and across the organisations that depend on it. Platform and migration decisions affect service continuity and local delivery. Data movement affects accessibility, privacy and trust. Reporting depends on consistent definitions and reliable flows. Clinical and operational improvement depends on information reaching the right people at the right time, without increasing administrative burden on frontline or data teams.

Delivery experience from within the estate demonstrates how closely these dependencies are already connected.

At platform level, critical workloads have been migrated from ageing private-cloud infrastructure into a modern enterprise environment across an estate of approximately **8PB of data, 900 datasets, 860 workloads and 1,000 users.** Our experience across major delivery programmes demonstrates that data platforms, pipelines, governance, reporting and outcomes services cannot be modernised independently. Decisions made in one area have material implications for delivery, resilience and value creation across the wider estate.

That work now connects directly into broader rationalisation across legacy and target environments. Current delivery includes estate inventory, dependency validation, controlled decommissioning, migration to FDP and NDIT, and bidirectional data movement between existing and future-state platforms. Reusable conversion blueprints and migration patterns are being established to make subsequent transitions more standardised and repeatable.

The same pattern is visible across operational data services. Critical pipelines have been assessed end-to-end, with source systems, transformation logic, validation processes and downstream dependencies mapped alongside strengthened monitoring, knowledge transfer and service management. A common privacy-treatment capability is being implemented across national datasets, with 70 datasets currently in scope and integration spanning multiple data platforms, cloud environments and downstream services.

Further downstream, work is underway to establish a more consistent architecture for metrics, lineage, governance and publication, while rationalising existing reporting services and dependencies across the wider analytical estate. In parallel, future-state outcomes capability is being designed around the integration of data sources, registries and clinical workflows to strengthen traceability, outcomes monitoring, safety surveillance and harm reduction.

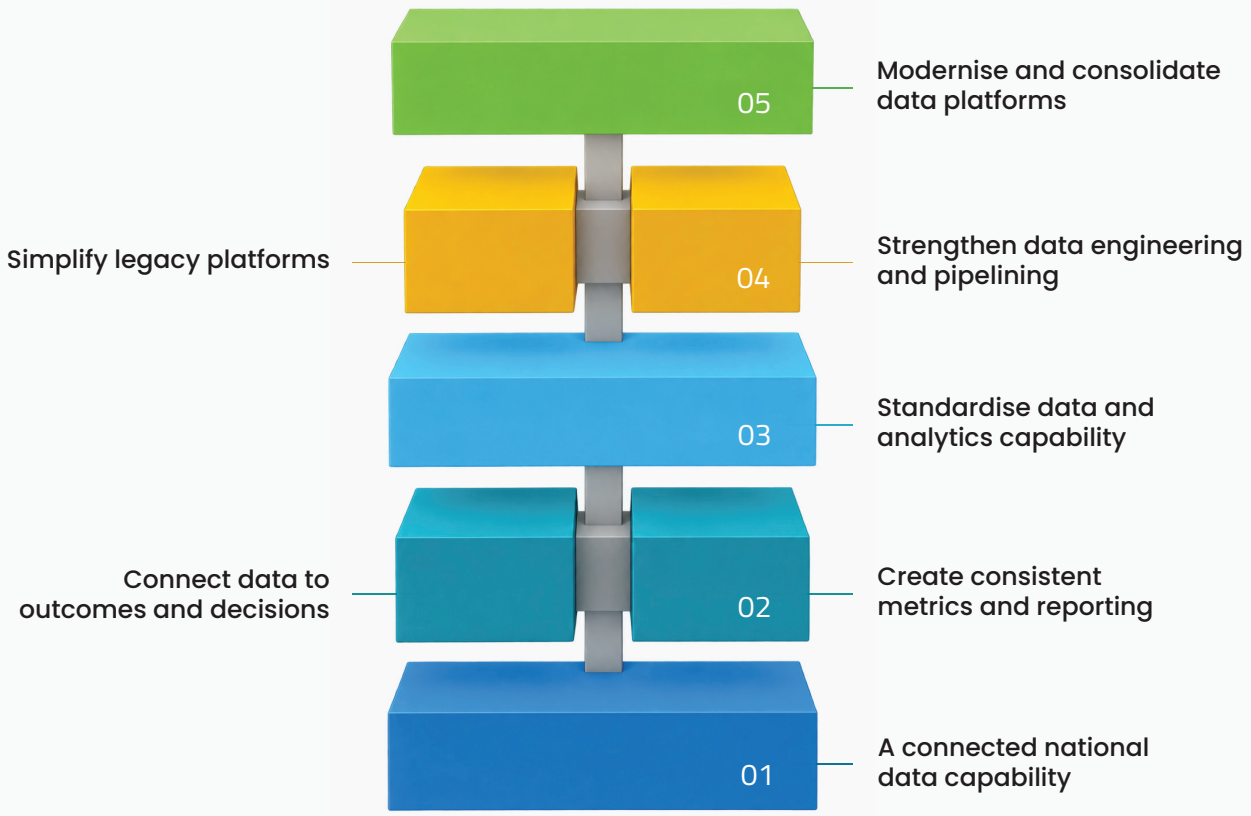
Taken together, this delivery experience points to a clear strategic requirement: **the data estate must be managed as an interconnected system, with dependencies governed across the transformation rather than resolved independently within individual workstreams.**

That has direct implications for delivery. Modernisation must preserve continuity across business-critical services that are central to the wider transformation. Migration should reduce technical debt and duplication as well as move workloads. Integration and governance patterns should be designed for reuse. Privacy, assurance and public trust should be treated as design requirements, with operational support increasingly built around shared capabilities. Delivery decisions should be made with a clear understanding of upstream and downstream impact.

The objective is an estate that becomes progressively simpler to operate, safer to change and easier to scale. Each phase of transformation should leave behind stronger operational resilience, clearer ownership and more reusable capability.

Success will ultimately be measured by the creation of a national data estate that functions as a coherent system: resilient, scalable and capable of converting trusted data into measurable clinical and operational improvement at pace.

Achieving operation excellence is already in flight with Mastek's Transformation Stack



The national data estate is being transformed from the foundations upwards

The individual components of a modern national health data capability are already being built across central and local services. They are best understood as a connected portfolio of capabilities rather than a single technology stack or linear delivery sequence. Some areas share common foundations in platforms, data engineering, privacy and governance, while others serve different users and objectives, including reporting and transparency, operational oversight, patient safety and outcomes improvement.

The value comes from greater consistency and coordination across those capabilities. Reliable data flows support both reporting and outcomes services. Shared privacy, governance and access controls reduce duplication. Modernised platforms and simpler legacy estates provide a more stable base for change. The aim is to allow different services to evolve against common standards and infrastructure, without forcing them into a single hierarchy or operating model.

Outcomes, registries and operational use cases

The underpinning infrastructure enables data to be turned into different forms of value across the system, from more transparent and consistent reporting, to better operational oversight, stronger accountability, improved patient safety and more effective use of outcomes data.

Current delivery is already working through what this requires in practice. Future-state national outcomes capability is being designed across clinical workflows, registry data, national data sources and external systems. The work has required integration feasibility assessment, minimum dataset definition, data lineage, prototyping and validation of some of the highest-risk assumptions before implementation.

This gives a direct view of what downstream services require from the underlying estate: dependable integration, clear lineage, usable data and architecture that can accommodate clinical and operational workflows rather than simply move information between systems.

Trusted metrics, reporting and insight

Immediately beneath the outcomes layer sits the need to turn data into a consistent version of reality.

Work currently underway across the reporting estate includes common metric definitions, versioning, lineage, suppression rules, publication controls and the architecture required to create a more authoritative metrics layer. At the same time, existing reporting pipelines have to remain operational while duplicated and resource-intensive processes are identified for rationalisation.

The practical insight is that reporting rationalisation reaches much further down the stack than dashboards and publications. It depends on the stability of upstream flows, common data models, privacy treatment and decisions about which existing services can safely be retired.

Privacy, governance and data enablement

Trust should be designed into the fabric of the estate rather than applied at its boundaries.

A common privacy-treatment capability is already being implemented across 70 in-scope datasets, spanning staging and production environments, data engineering platforms and downstream consumers. Delivery includes treatment logic, derivation tables, production testing, monitoring, customer migration and knowledge transfer, creating a more consistent and scalable approach to protecting sensitive information across national services.

This experience highlights an important characteristic of the future-state architecture: privacy, governance and assurance are increasingly becoming enabling capabilities rather than compliance controls alone. As data moves across platforms, organisations and analytical services, trust must be established once and reused many times. Rebuilding governance, assurance and privacy controls for individual data products introduces cost, delay and inconsistency, while increasing operational risk.

Shared controls, common standards and reusable assurance patterns allow new services to inherit trust rather than recreate it. This reduces delivery friction, accelerates onboarding to new platforms, supports consistent regulatory compliance and provides greater confidence in the quality, lineage and appropriate use of data. As the estate evolves, privacy, governance and assurance should increasingly function as foundational infrastructure: supporting innovation, enabling safe data sharing and providing a trusted basis for clinical, operational and analytical decision-making at scale.

Data pipelines, integration and legacy rationalisation

Below these capabilities sits the most operationally complex part of the estate.

Current work has mapped critical data journeys from source through transformation, validation, aggregation and consumption, exposing business rules, integration points, manual interventions and downstream dependencies that are often difficult to see from platform architecture alone.

At the same time, rationalisation activity is working across legacy data platforms, SQL environments and newer target environments. This includes controlled decommissioning and bidirectional movement between legacy services, FDP, NDIT and secure data environments.

This is where much of the real complexity of transformation sits. The estate cannot simply be redrawn from the top down. Existing flows, dependencies and operational behaviour have to be understood well enough to determine what should move, what should change and what can disappear.

Modern platforms and infrastructure

The foundation is the continued modernisation of the underlying platform estate.

Large-scale migration has already been undertaken across approximately **8PB of data, 900 datasets, 860 workloads and 1,000 users,** requiring discovery, dependency analysis, migration patterns, automated reconciliation, controlled deployment and operational transition.

The value of that foundation increases as the layers above it are simplified with it. Modern infrastructure creates the conditions for change; the wider return comes from using that change to reduce legacy dependencies, standardise data movement and create reusable capabilities across the estate.

The picture that emerges is of one connected transformation. The layers are at different stages of maturity, but they already intersect through common datasets, platforms, engineering patterns, controls and operational dependencies.

The depth of the opportunity lies in those interfaces: understanding how a change at one layer affects the services above it and the infrastructure below it.

Four principles for delivering operational excellence

Principle 1

Operate while you transform

Large-scale data transformation takes place inside a live operational environment. National services, reporting flows and analytical products continue to support day-to-day decision-making while platforms, pipelines and operating models are changing around them.
That makes continuity a design requirement rather than a transition activity.

Recent delivery across nationally critical data services has shown the value of combining transformation with detailed dependency mapping, knowledge capture, monitoring, support-model design and operational risk assessment. In one major data-engineering engagement, source systems, transformation logic, validation processes and downstream dependencies were mapped end-to-end to strengthen resilience and reduce dependence on undocumented operational knowledge.

The same discipline applies at platform scale. Migration of approximately 860 workloads, 900 datasets and 8PB of data required structured discovery, automated validation, business readiness, controlled deployment and hypercare alongside the technology transition itself.

The principle is straightforward: **The migration path is as important as the target architecture.**

Structured knowledge capture, robust operational resilience and transition methodologies, service ownership and transition planning should therefore be designed into transformation from the outset.

Principle 2

Remove complexity as part of migration

Modernisation creates limited value if existing complexity is simply recreated on newer infrastructure.

Migration should be used as an opportunity to challenge redundant services, duplicate pipelines, unnecessary infrastructure and legacy dependencies. This requires a disciplined understanding of the estate before decisions are made about what should move, what should be redesigned and what should be retired.

Current rationalisation activity across multiple data environments is applying this approach through estate inventory, dependency validation, controlled decommissioning, workload migration and the development of reusable conversion patterns.

The value compounds over time. Reusable patterns for discovery, testing, assurance and decommissioning reduce the need to redesign each migration independently and create a more consistent route through the estate.

Every migration should leave the estate simpler, reduce the burden on those supplying and managing data, and make the next transition easier to execute. Where data quality and control can be maintained, federated access should progressively replace unnecessary central collection.[sf13.1]

This provides a stronger basis for long-term cost reduction and operational resilience than workload movement alone.

Principle 3

Treat interoperability and trust as shared capabilities

A modern health data estate will remain heterogeneous. Multiple platforms, cloud environments, legacy services and analytical tools will continue to coexist for the foreseeable future.

The critical capability is therefore the ability to move data between them safely, consistently and with clear governance.

Existing delivery demonstrates the importance of reusable integration patterns. Bidirectional data movement has been established across legacy environments, modern platforms, SQL services and secure data environments, supported by common orchestration and migration mechanisms.

Trust must be designed with the same level of reuse. A common privacy-treatment capability is being implemented across national datasets, supported by standardised treatment policies, derivation logic, testing, monitoring and migration controls.

The strategic direction is towards shared services for integration, privacy, governance and assurance that can be consumed across multiple use cases.

This reduces duplication while improving consistency and control across the estate.

Principle 4

Maintain a clear line of sight from data to outcomes

Technology transformation creates value through the decisions, services and outcomes it enables.

That requires stronger control over the journey from underlying data to the information used by operational leaders, clinicians and analysts. Metric definitions, lineage, governance and publication controls become increasingly important as reporting is consolidated and reused across multiple services.

Current work to establish a central metrics capability is addressing these requirements through common definitions, versioning, lineage, governance and controlled publication, alongside the rationalisation of existing reporting activity.

The same logic extends into clinical outcomes. National registry and outcomes capability depends on the effective integration of data sources, clinical workflows and analytical services to support traceability, safety surveillance and outcomes monitoring.

Public trust is a further dependency. Greater use of linked data, analytics and outcomes information will only be sustainable if governance, transparency and patient choice are built in from the outset. This requires clear involvement of relevant data governance bodies, transparent public communication on how data is used, and continued adherence to established opt-out mechanisms.

This creates a clear value chain: **Data source -> governed flow -> trusted information -> decision -> outcome**

Measures such as workloads migrated, pipelines automated and platforms retired remain important. They need to sit alongside measures of resilience, efficiency, decision quality, service performance, clinical value and public confidence.

What could derail the transformation

The hardest risks are buried in the existing estate

The main delivery risks are not in the target architecture, but rather in the legacy estate: anywhere where there is undocumented logic, hidden dependencies, manual interventions, duplicated flows and knowledge held by individual people.

Those are the issues that determine whether transformation actually reduces complexity.

Moving workloads without understanding what they carry

A pipeline is rarely just a pipeline. It can contain business rules, validation logic, downstream dependencies and manual fixes that have built up over years.

That has been visible in current delivery. End-to-end analysis of critical data flows has had to trace source systems, transformations, validation steps and downstream consumers before changes could be made safely.

The same is true of decommissioning. Retiring a service has required dependency checks, rollback plans, archival controls and cold-storage migration before it could be switched off.

The practical lesson is simple: moving a workload is often easier than proving the old one can disappear.

Assuming the move from legacy to target state will be one-way

It will not be one-way. Current delivery already depends on two-way movement between legacy platforms, FDP, NDIT, SQL environments and secure data environments, including write-back into older services.

That means coexistence has to be designed properly. For a significant period, old and new environments will both remain part of the operating landscape.

If that period is poorly controlled, the result is duplicate processing, unclear ownership and multiple versions of the same data flow.

Losing the people who know how the estate actually works

Some of the most important knowledge still sits with individuals.

Work on critical pipelines has exposed gaps in runbooks, undocumented processing steps and areas where operational knowledge was concentrated in a small number of specialists.

Addressing that has required structured knowledge capture, validation of operational documentation and deliberate transfer into support teams.

That becomes more important during supplier transition, organisational change and platform migration. If knowledge transfer happens at the end, it is already too late. Operational knowledge needs to move with the service.

Optimising one part of the estate at the expense of another

The delivery environment is heavily interdependent. Platform vendors, engineering teams, privacy services, run-and-maintain teams and specialist suppliers are often working across the same datasets and services.

That means a sensible local decision can still create problems elsewhere.

A privacy change can trigger customer migration. A reporting change can depend on upstream metric design. A decommissioning decision can be blocked by consumers outside the immediate workstream.

This is where governance often becomes too programme-specific. Milestones can be green while cross-estate dependencies remain unresolved.

The more useful test is whether each change removes something from the estate: a dependency, a manual process, a duplicate service, a fragile hand-off, a piece of undocumented knowledge.

If it does not, complexity has probably just moved.

A virtuous data cycle for the national data estate

Creating a reinforcing cycle of better data, better decisions and better outcomes

THE COMPOUNDING VALUE LOGIC

Strategic compounding

Each turn of the flywheel builds momentum, making the next change easier and less costly to execute.

EACH CHANGE SHOULD REDUCE THE COST, RISK OR EFFORT OF THE NEXT ONE

Build once, reuse, improve continuously

A large and complex data estate becomes easier to change only when each phase of transformation leaves behind something reusable.

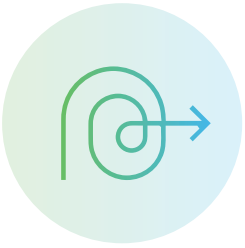
That means moving away from one-off delivery patterns and towards a model where knowledge, controls and technical approaches are progressively standardised across the estate.

Understand

Create a reliable view of the estate before making change: critical services, dependencies, business rules, users, operational risks and downstream consumers.

Current delivery has shown why this matters. Critical pipelines have required end-to-end mapping of source systems, transformations, validation logic and downstream dependencies before transition decisions could be made safely.

Simplify



Use transformation to remove redundant services, duplicate flows and unnecessary infrastructure.

Current rationalisation work combines estate inventory, dependency validation, controlled decommissioning and migration. This prevents legacy complexity being carried forward into newer environments.

Standardise

Turn proven approaches into common patterns. This is the mechanism that produces the standardised feedback loops and shared playbooks as per NHS delivery models.

Reusable migration blueprints, conversion patterns, validation methods and governance controls are already being developed across the estate. Standardisation reduces delivery variance and makes future change more predictable.

Automate

Reduce the number of manual steps required to operate and change the estate.

Automated reconciliation, validation and testing have already been used to support large-scale workload migration. The same principle applies to monitoring, data movement, privacy treatment and reporting.

Productise

Treat recurring capabilities as reusable services.

Privacy treatment, data movement, migration tooling, metric management and publication controls all have the potential to be consumed repeatedly across different use cases rather than rebuilt each time. Current delivery is already establishing common treatment logic, derivation capability and operational handover across multiple datasets and environments.

Measure

Track whether the estate is actually becoming easier to operate.

Useful measures extend beyond delivery volume. They include reduction in legacy infrastructure, fewer duplicated services, lower dependency concentration, increased automation, faster transition cycles a stronger service resilience, and improved decision-making and outcomes.

Improve

Feed delivery evidence back into the model.

Patterns should change as teams learn which approaches work, where risk concentrates and which controls slow or accelerate delivery. The objective is cumulative improvement rather than repeated reinvention.

Each change should reduce the cost, risk or effort of the next one.

That is the discipline required to turn a collection of modernisation activities into a more coherent national data capability.

The opportunity

The foundations of a modern national health data capability are already in place. The next phase is to make those foundations work together more consistently, with less duplication, less operational friction and greater reuse across the estate.

A mature data estate should become easier to operate and easier to change over time.

Platforms should be simpler. Data movement should rely on more consistent patterns. Privacy and governance should be embedded into shared capabilities. Metrics should be trusted and reusable. Operational knowledge should be retained as services evolve.

Delivery experience across platform migration, legacy rationalisation, critical data pipelines, privacy treatment, reporting architecture and outcomes capability shows that this is achievable.

It also shows that the greatest value is created at the interfaces between these disciplines: where migration meets continuity, where data movement meets privacy, where reporting meets common definitions and where infrastructure decisions enable [KE17.1]downstream outcomes.

The strategic opportunity is to turn a series of transformation activities into a coherent national capability that reduces complexity as it evolves, protects critical services during change and creates a more direct route from trusted data to operational and clinical value.

The measure of progress is simple: the estate should become easier to change with every change made.

Mastek
Trust. Value. Velocity
Lead with AI

About us

Mastek (NSE: MASTEK; BSE: 523704) is a global provider of enterprise AI, digital, and cloud services, enabling clients to achieve measurable and sustainable returns on their technology investments. The company has a presence in over 40 countries and a skilled workforce of close to 5,000 employees. Through its "Lead with AI" approach, Mastek integrates intelligence across its solutions and operations, enabling organizations to accelerate transformation using ethical, scalable, and domain-driven AI adoption. Mastek partners with industry leaders such as Oracle, Salesforce, Microsoft, AWS, Snowflake, and Databricks, serving key sectors such as Public Sector, Healthcare, Retail, Manufacturing, Higher Education, and Financial Services. Mastek is committed to driving innovation by developing a strong ecosystem of start-ups, academia, and IPs. With its core values of trust, value, and velocity, the company empowers 400+ active customers to transform their business in the evolving tech landscape. Mastek has always been a solutions focused and relationship-centric company, valuing both employees and clients. With its humane approach, Mastek fosters growth through sustainable goals, high ethical standards, and responsible governance.

Contact us:

info@mastek.com www.mastek.com