

FEBRUARY 2025
HENRY LI
ANASTASIA
BEKTIMIROVA
AXEL HEITMUELLER
DARCY WARD
ALEXANDER IOSAD



Governing in the Age of AI: Building Britain's National Data Library

Contents

- 3 Foreword
- 6 Executive Summary
- 11 Introduction
- 26 The Road to a National Data Library
- 58 Conclusion
- 59 Acknowledgements

Contributors: Robert Johnson, Charlotte Refsum

***Governing in the Age of AI: Building Britain's National Data Library* is a joint report by the Tony Blair Institute for Global Change and [The Entrepreneurs Network](#).**

Foreword

The United Kingdom should lead the world in artificial-intelligence-driven innovation, research and data-enabled public services. It has the data, the institutions and the expertise to set the global standard. But without the right infrastructure, these advantages are being wasted.

The UK's data infrastructure, like that of every nation, is built around outdated assumptions about how data create value. It is fragmented and unfit for purpose. Public-sector data are locked in silos, access is slow and inconsistent, and there is no system to connect and use these data effectively, or any framework for deciding what additional data would be most valuable to collect given AI's capabilities.

As a result, research is stalled, AI adoption is held back, and the government struggles to plan services, target support and respond to emerging challenges. This affects everything from developing new treatments to improving transport, tackling crime and ensuring economic policies help those who need them. While some countries are making progress in treating existing data as strategic assets, none have truly reimagined data infrastructure for an AI-enabled future.

No nation today has the infrastructure needed to fully harness AI for public good. The National Data Library (NDL) represents an opportunity for the UK to be the first. It can help create the infrastructure needed to unlock the value of public-sector data alongside frameworks to identify and collect new types of data for breakthrough insights. Instead of a complicated web of systems and slow, one-off approvals, the NDL will establish a clear, secure way to access linked data sets, supporting AI innovation, better

policymaking and research that drives economic growth. It will not centralise data but will put in place the legal, technical and governance structures to ensure that high-value data sets can be used efficiently while maintaining security and trust.

This can only work if the NDL is driven by vision, not just technology. Too many past government data initiatives have focused on systems and processes rather than real-world outcomes. The NDL must not fall into the same trap. It must be built to remove systemic barriers to data use, support AI-driven discovery and help us understand what additional data would be most valuable to gather given AI's analytical capabilities, delivering tangible economic and societal benefits. A narrow, technical approach will fail.

This report sets out a clear roadmap to get this right and make it work. It details the immediate steps to unlock more value from existing data, the structural reforms needed to build a scalable system and the long-term actions necessary to ensure the NDL delivers sustained impact by reimagining what data we need in an AI-enabled world. The government must take this seriously, act at pace and use this report as a blueprint for delivery.

Nathan Benaich, Founding Partner, Air Street Capital

Lord James Bethell, former Minister for Life Sciences

Dr Sarah Cumbers, CEO, Royal Statistical Society

Priya Guha MBE, Non-Executive Director and Advisor

Barney Hussey-Yeo, Founder and CEO, Cleo

Kirsty Innes, Director of Technology Policy, Labour Together

Nicklas Berild Lundblad, Senior Director of Policy and Strategic Advisor,
Google DeepMind

Lord James O'Shaughnessy, former Minister for Innovation

Sam Roseveare, Director of Regional and National Policy, University of Warwick

Lord David Willetts, President, Resolution Foundation; former Science Minister

Julia Wilson, Director of Strategy, Partnerships and Innovation, Wellcome Sanger Institute

The support expressed by the signatories above is their own and does not necessarily reflect the views of their employer.

02

Executive Summary

The National Data Library (NDL) has the potential to become a vital piece of enabling infrastructure for public-service delivery and economic growth in the United Kingdom. It is intended to unlock the full potential of public-sector data by enabling secure, seamless, quick and scalable access to linked data sets. By breaking down barriers to data use, it can drive artificial-intelligence-enabled innovation, transform public services and fuel the development of solutions to national challenges that are uniquely tailored to the UK context.

There is considerable excitement around the idea of the NDL, with a vibrant debate around its objectives and structure in Parliament, among researchers and in civil-society organisations. However, there is still no clear, bold vision for its purpose or how it should be delivered. To succeed, **the NDL must be vision-led, not technology-led**. Its development should be driven by impactful use cases, aligned with national priorities and structured to quickly begin to deliver tangible benefits to citizens, researchers and businesses alike.

There are currently competing proposals for the NDL, reflecting different perspectives across government and expert groups. While these proposals are not mutually exclusive, it is important to be clear-eyed about where government is well placed to add distinctive value and break down systemic barriers hampering progress, and where the private sector, philanthropic initiatives and research funders are better positioned to lead. The guiding questions should be these: What is the NDL's unique value-add? And what can only be achieved with the NDL as vital enabling infrastructure?

The NDL's transformative potential lies in its ability to advance better decision-making, support innovation and drive AI-powered applications across multiple domains. It is not just a data-sharing platform; it is a strategic enabler of evidence-based policymaking, research and economic growth. It is essential to delivering the government's Plan for Change, and supporting key missions including improving health care, tackling inequality, boosting economic growth and achieving clean-energy goals.

For example, integrating National Health Service and Department for Work and Pensions (DWP) data sets could enable tailored return-to-work programmes, improving productivity and public health. Linking planning, environmental and energy data could accelerate housing and green-energy solutions, while integrating education data sets could help identify links between pupils' assessments and economic mobility, shaping more effective teaching strategies.

The NDL can amplify the impact of linked data across government missions, delivering measurable societal and economic benefits that improve lives and strengthen communities. Analysis later in this report (see The Economic Case for the NDL) estimates that an initial investment of £200 million to enhance data linking could generate £1 billion in benefits in the first instance. As the NDL expands, reinvesting these returns could create a self-sustaining model, delivering up to £13 billion per year once fully scaled.

To realise this vision, the NDL must avoid the pitfalls of past large-scale government information-technology projects – which have often lacked agility, professional and public legitimacy, and a clear delivery pathway. The government must act with urgency and clarity, ensuring the NDL is delivered at pace and with clear accountability. The prime minister must take charge, aligning the NDL with the objectives of the AI Opportunities Action Plan. This will ensure that the UK positions itself as a global leader in fostering AI-driven innovation and economic growth, as well as an exemplar of best practice in using AI to reimagine the state.

The competition is fierce, with the United States and China dominating general-purpose AI. But as the action plan sets out, the UK's competitive edge lies in being “an AI maker, not an AI taker” through the development of specialised models and applications where its data advantages are strongest. The NDL is central to the UK's ability to go from competing for bronze in general-purpose AI to securing gold in domains that will be critical to economic growth and disruptive delivery.

In this report, the Tony Blair Institute for Global Change, in collaboration with The Entrepreneurs Network, sets out a roadmap for building the NDL that is structured across three key phases:

Immediate actions (first six months): Unlock value from existing data, with an early focus on existing administrative data processes, by streamlining access processes, demonstrating early results and building momentum. This requires accelerating the process of receiving approvals for sharing data, supporting data controllers in preparing data sets for linkage and addressing fragmentation in information-governance frameworks. Public trust and accountability must be embedded in the NDL's governance from the outset, to avoid the pitfalls of insufficient transparency and public engagement that impacted past data initiatives.

Medium-term actions (six months to three years): Scale the NDL's infrastructure and governance to support broader impact. A federated data-sharing model should be implemented, ensuring that data sets remain at their source while being securely linked. Pre-linked data sets should be developed for high-value applications such as health-employment integration, enabling AI-driven solutions for public services, with strategic assets such as address data made available more widely through the NDL. Access to data sets should be coupled with access to compute, further cementing the NDL's contribution to economic growth. Governance structures must be strengthened, including the enactment of legal changes to expand access for responsible commercial use. The NDL must be put on a statutory footing through a new Act of Parliament to secure its political independence as vital enabling infrastructure.

Long-term actions (three to five years): Realise the full vision of the NDL as a unified, scalable platform that transforms how the UK uses data. Identifiable data should be integrated under strict safeguards, enabling applications such as personalised health care, targeted social interventions and precision policymaking. Harmonised personal identifiers, using a consistent number to refer to the same entity in different places, should be introduced to improve interoperability and linkage across government systems. The NDL should also establish Data Biomes – collaborative hubs where government, academia and industry tackle complex challenges together. These biomes should align with key national priorities such as climate resilience, public health and economic growth, ensuring continuous innovation and targeted, data-driven solutions.

The government must act now and build towards an ambitious vision. The roadmap is clear, the framework is in place and the potential benefits are transformative. The NDL will position the UK as a leader in data-driven innovation while ensuring public-sector data deliver real-world value for society, the economy and AI-driven breakthroughs. Now is the time to build.

Key Policy Steps Across the Phases

IMMEDIATE ACTIONS (FIRST SIX MONTHS)

- **Establish a senior leadership team** to oversee implementation, with a mechanism for receiving regular input from and providing updates to ministers from key user departments including the Treasury and Cabinet Office.
- **Unlock value from existing data sets** by streamlining approvals, supporting data controllers and addressing fragmentation in data-sharing frameworks.
- **Launch a network of National Data Librarians** embedded in key departments and relevant public-sector bodies to improve data quality, support data integration and drive early adoption.
- **Develop a public transparency framework**, including a registry of approved projects, data sets accessed and their intended use.
- **Build a strong business case for long-term investment**, ensuring the NDL's financial sustainability beyond the initial phase.

MEDIUM-TERM ACTIONS (SIX MONTHS TO THREE YEARS)

- **Implement a federated data-sharing model**, enabling data sets to remain at their source while being securely linked.
- **Develop pre-linked data sets** for high-value applications such as health-employment integration to support AI-driven public-service improvements.
- **Expand analytical and compute capabilities**, ensuring AI-readiness through integration with the AI Research Resource.

- **Introduce an NDL Reader Pass with a tiered access system**, balancing security with usability to enable streamlined access for trusted users. Introduce a Data Offenders Register for those found to abuse this trust.
- **Develop a sustainable commercial model**, ensuring financial viability through tiered data-access fees while maintaining accessibility for research and innovation.
- **Introduce and pass the NDL Bill**, putting the NDL on a statutory footing as an arm's-length body to ensure political independence and long-term stability.

LONG-TERM ACTIONS (THREE TO FIVE YEARS)

- **Integrate identifiable data under strict safeguards**, enabling applications such as personalised health care, targeted social interventions and precision policymaking.
- **Introduce harmonised personal identifiers** to improve data linkage across government systems, increasing efficiency and interoperability.
- **Expand Data Biomes** as collaborative hubs for government, academia and industry to tackle national challenges such as climate resilience and economic growth.
- **Scale international engagement**, positioning the NDL as a global leader by integrating it into the UK's AI and data-driven innovation strategy.

03

Introduction

Public services in the UK face a fundamental paradox: they are data-rich but insight-poor – even more so when it comes to actionable insights. While vast amounts of data are collected across government departments, much of these data remain fragmented, difficult to access and underutilised. Even when linked, – that is, when records that refer to related entities but reside in different places are connected, for example, through a common identifier – data sets are not consistently applied to inform policy, improve public services or drive innovation. This lack of integration and use prevents the government from making faster, smarter decisions, hampers research and development, and constrains economic growth.

The establishment of a National Data Library (NDL) is an opportunity to fundamentally change this dynamic. The NDL could unlock the full potential of public-sector data by making it easier, faster and safer for government, academia and industry to access and use them. Unlike past efforts – such as the Office for National Statistics (ONS) Secure Research Service, which has improved controlled access but remains limited in scope, and the Integrated Data Service, which enhances linkage but is not yet a fully scalable solution – the NDL should not just be about making data available; it should also be about making data work hard.

For the government, the NDL would enable better policymaking, real-time operational foresight and more effective public services. For researchers, it would accelerate breakthroughs across disciplines – from life sciences and social sciences to cutting-edge AI research. For industry, it would provide a platform for innovation that powers economic growth in the UK and enables businesses to develop tailored solutions to societal and national challenges.

Recognising the urgency of this need, the government has made the NDL a central pillar of its transformation agenda. Since the 2024 general election, when Labour's manifesto included a commitment to establishing the NDL, ministers have repeatedly emphasised its importance. It has been referenced 29 times in Parliament, through responses to written questions and speeches in both the Commons and the Lords, and has featured in five

government policy papers. The NDL is set to be a cornerstone of the government's efforts to deliver its five cross-departmental missions: driving economic growth, achieving its clean-energy goals, tackling inequality, transforming health care and strengthening public safety.

The AI Opportunities Action Plan¹ positions the NDL not just as a way to improve data access but as a strategic enabler, powering the development of advanced AI systems and supporting real-world applications across a wide range of domains.

However, with such a broad range of expected benefits, there is a risk that the NDL's core purpose could become diluted. Clarity of focus is required, both for its purpose and for its scope.

The NDL's main purpose should be to remove systemic barriers to data access, ensuring that high-quality, AI-ready public-sector data can be safely and efficiently used by scientists and private-sector innovators, while enhancing public services and maintaining public trust. This vision of turning fragmented public-sector data into a powerful, accessible resource was set out in UK think tank Onward's original "British Library for Data" proposal as a stride towards the UK's strategic advantage, and led to the NDL being formalised as a commitment in Labour's growth-oriented manifesto and subsequent policy framework.²

As for its scope, the NDL should in the first instance focus on access to administrative data. Here, the government has unique capabilities and clear responsibilities. Administrative data – generated through routine government operations such as tax records, welfare payments, school enrolments, health-service usage, weather records and road-traffic data – are distinct from other data sources. Unlike survey data or commercially collected data sets, they provide a comprehensive, near-real-time view of societal trends and public-service outcomes.

While data-driven innovation spans both public and private sectors, industry-led and open data sets already benefit or can benefit from dedicated initiatives such as Smart Data schemes.³ In contrast, government-held data is fragmented, lacks a cohesive system for access

and use, and so remains underutilised. The NDL's role should not be to duplicate commercial or research efforts but to unlock the value of these data by addressing these structural, technical and legal barriers.

Improved public-sector data-set access could unlock billions in economic returns. Even simply linking these could generate at least £1 billion in returns, based on Administrative Data Research (ADR) UK's demonstrated impact.⁴ Over the long term, wider societal benefits from improved data access and research applications could reach £319 billion by 2050 – equivalent to £13 billion per year.⁵ Further, a fully developed AI-ready data ecosystem could drive a 20-fold return on investment, with high-end estimates reaching £199.7 billion over five years, or £40 billion annually in productivity gains.⁶ The NDL is essential to realising this potential and ensuring the UK maximises the value of its public-sector data.

The NDL should become the service layer for secure, structured and real-time access to linked public-sector data sets. Rather than centralising data, it would provide the infrastructure, governance and access controls needed for responsible, efficient and scalable data sharing. By establishing a trusted framework, enforcing security and compliance, and facilitating the federation of linked data sets, the NDL would ensure that government-held information remains decentralised but accessible when needed.

None of this would be possible without efforts to improve the broader data infrastructure, including efforts around interoperability and digital identity. Thankfully, the Digital Centre for Government Blueprint, recently published by the Department for Science, Innovation and Technology (DSIT), lays out an ambitious programme of work that, if delivered at pace, would create the necessary conditions.⁷

This allows the NDL to focus on closing a critical gap by addressing the legal, operational and structural barriers that prevent effective data use. Interoperability and even linkage efforts, welcome as they are, do not guarantee access or usability. Data sets must be structured to ensure they can be efficiently combined and queried across different systems. This principle has been central to recent work on the NDL's design, including insights from the Wellcome and Economic and Social Research Council

(ESRC) Technical White Paper Challenge on the technical foundations of the Library.⁸ There is broad agreement that the NDL should not attempt to centralise all data but instead serve as the connective layer that links existing infrastructure through federated service solutions.

Complementing this technical foundation, this report provides a clear roadmap for delivery – moving beyond design principles to outline how the NDL can be built, scaled and embedded as a lasting feature of the UK's data ecosystem. We set out immediate actions to unlock early benefits, medium-term steps to consolidate progress and long-term strategies to ensure the NDL delivers sustained impact. The challenge is not just to design a system but to ensure it works in practice, driving measurable improvements across government, research and industry from day one while working towards a shared greater ambition.

The NDL must avoid three key pitfalls: becoming bogged down in grand-but-distant visions that fail to gain trust, fragmenting into a patchwork of disconnected IT projects, or losing sight of its unique value by investing in ideas that could have been realised without it.

To succeed, the NDL must be vision-led, not technology-led. It should measure its success by whether researchers are able to rapidly access linked data sets, innovators can develop AI-driven solutions, policymakers can make informed, data-driven decisions in real time, and data controllers – the organisations that hold data, and decide why and how they may be used by others – have the tools to share data securely and effectively.

What Is the National Data Library?

Fragmented systems, inconsistent data-sharing frameworks and a lack of coordination continue to impede access and limit impact of public-sector data assets. This challenge is particularly urgent as the AI Opportunities Action Plan presents a window for AI to transform public services, science and economic growth. However, AI is only as powerful as the data it learns from, and much of the UK's most valuable public-sector data remain siloed, inaccessible and poorly structured for AI training. The NDL could change this

by linking, standardising and securing key data sets, ensuring AI has the real-world, high-quality data needed to drive smarter policymaking, faster research and more effective innovation.

Rather than acting as another government IT project, the NDL should give government departments the tools, expertise and governance needed to improve how data are prepared, shared and used. This distinctive role is central to its purpose.

THE NDL'S ROLE

What It Is and What It Is Not

THE NDL SHOULD BE:

- **A centre of excellence**, equipping departments with the tools, standards and expertise to improve data quality, facilitate responsible data sharing and unlock new opportunities for collaboration.
- **A catalyst for impact**, actively enabling high-value data integration that supports government missions, research breakthroughs and economic growth – delivering measurable benefits across public services, academia and industry.
- **A secure environment** for discovering, accessing and using linked public-sector data across government departments, ensuring data are available for research, policymaking and innovation while maintaining the highest privacy and security standards, with strong disincentives for data misuse.

THE NDL SHOULD NOT BE:

- **A giant data lake** that centralises all government-held data in one place. Instead, it would enable secure and federated access while ensuring departments retain control over their own data.
- **A marketplace for selling government data** – the NDL would not be a commercial data broker. While it should develop a sustainable access model for third parties that recognises the commercial value of these data, its purpose is to facilitate their responsible and ethical use, not to monetise public-sector information.

- **A superficial branding exercise** – the NDL cannot simply be used to re-label existing data-related initiatives. The NDL's unique value should be realised by building the infrastructure, governance and services needed to make data more accessible, useful and impactful in ways that were previously impossible.

THE NDL WILL BE A SUCCESS IF IT:

- **Transforms data access** by making it easier, faster and more secure to link and use government data.
- **Standardises and simplifies** data-sharing processes, removing bureaucratic obstacles while ensuring compliance with legal and ethical standards.
- **Empowers data controllers** with the support and guidance needed to prepare data for responsible use, rather than simply demanding access.
- **Delivers real-world impact** by enabling high-value use cases that improve public services, inform evidence-based policymaking and drive growth through innovation.

To achieve its vision, the NDL must build on existing initiatives while addressing persistent barriers to effective data use. The ADR UK programme has already committed £105 million to improving government data access for research, including support for the ONS Secure Research Service, which provides approved researchers with secure access to administrative data sets.⁹

Building on this foundation, the ONS Integrated Data Service (IDS) enhances the ability to safely link and analyse de-identified public-sector data. It currently hosts more than 100 data sets, including linked census, tax and health records, with plans to incorporate benefits, justice and education data, further increasing its utility.

These efforts provide a strong starting point for the NDL, enabling it to leverage existing data sets and infrastructure. However, challenges remain in accelerating data access, broadening usability across government, academia and industry, and supporting data controllers in preparing data sets for linkage. To fully unlock public-sector data's value, the NDL must establish itself as a service layer, not just a data repository.

The NDL's transformative potential lies in its ability to unlock high-impact applications across four interconnected domains: academic research, commercial research and development (R&D), policy design and public-service delivery. Each relies on linked data to drive meaningful outcomes, and progress in one area can amplify benefits across others, directly supporting the government's core missions (see Figure 1).

FIGURE 1

Potential use cases of the NDL, once fully developed, and their role in delivering government missions

Kickstart economic growth

Develop predictive models for economic forecasting by linking HMRC, DWP and Treasury data.

Reduce inactivity by providing tailored and personalised support to specific groups via DWP/NHS data.

Reduce construction costs for house building by leveraging the National Underground Asset Register (NUAR).

Attract life-sciences investment by creating linked, longitudinal health records.

Speed up planning approval with a database of planning documents and local economic plans that link data across planning authorities

Make the UK a clean-energy superpower

Identify suitable locations for housing by linking data from Ministry of Housing, Communities and Local Government and Department for Environment, Food and Rural Affairs (Defra) to avoid high-pollution areas.

Assess water stresses and flood risks by linking Department for Energy Security and Net Zero and Defra data sets and satellite imagery.

Distinguish green claims from greenwashing by linking Defra, financial- and economic-regulator data.

Create a comprehensive energy-use model by combining data from smart meters, the National Grid and local energy providers.

Safer streets

Improve the impact of probation programmes on employment prospects by linking DWP and Ministry of Justice data.

Improve threat assessments by linking national-security intelligence across the Home Office, Ministry of Defence and allied agencies.

Optimise the allocation of investment into the prison estate to support the right balance of offender management in the community and imprisonment for serious and repeat offenders

Break down barriers to opportunity

Link education, employment and social-care data to identify factors strongly correlated with social mobility.

Support people to develop the skills they need to grow in their careers by combining models of local labour-market development with education data.

Inform policies that address barriers to higher and further education and ensure provision meets current and future demand.

An NHS fit for the future

Link air-quality and health data to target respiratory diseases.

Facilitate better clinical trials by linking primary-care and secondary-care data.

Link health and socioeconomic data to target health inequalities.

Source: TBI

THE ECONOMIC CASE FOR THE NDL: UNLOCKING HIGH-RETURN INVESTMENTS

The NDL is not just a strategic data initiative – it is an economic opportunity. Evidence from existing data-linking projects demonstrates that investing in better data access delivers substantial returns. The ADR UK programme has already shown a 1:5 cost:benefit return on investment, and similar projects across government have generated significant cost savings and efficiency gains.¹⁰

If the government prioritised just ten high-impact use cases, the initial costs could be up to £200 million – a figure in line with recent government data initiatives such as the £20 million investment in linked data sets to improve border flows¹¹ and the Covid-19 data store.¹² With these targeted investments, the returns could exceed £1 billion simply by making existing data sets more accessible and usable.

In the longer term, fully modernising outdated government IT systems and expanding AI-driven innovation could unlock far greater value. Data and Analytics Research Environments (DARE) UK estimates that better data access could generate £319 billion in societal benefits by 2050 – equivalent to £13 billion annually.¹³ Further analysis from the Tony Blair Institute for Global Change underscores the transformative potential of AI-ready data ecosystems. With the right data infrastructure, AI-driven productivity gains could deliver a 20-fold return on investment in productivity, with high-end estimates reaching £199.7 billion over five years – or £40 billion per year.¹⁴ Even modest projections indicate significant economic benefits, making investment in the NDL a financially sound and necessary step for the UK's economic and technological future.

The following sections show how the NDL could turn this vision into reality – demonstrating its impact across four key areas: advancing research, accelerating AI innovation in the private sector, informing policy and transforming public services.

HOW COULD THE NATIONAL DATA LIBRARY BE USED?

Use Case 1 – Academic Research: Accelerating Discovery and Economic Growth

The NDL will provide the missing infrastructure for UK research, ensuring faster access to high-quality public-sector data and accelerating discoveries in medicine, social policy and climate science. Instead of constantly coming up against the limits of fragmented data sets and slow, inconsistent access practices, researchers will be able to generate insights at scale, accelerating breakthroughs that directly improve lives.

Medical scientists would be able to determine treatment effectiveness faster by linking NHS and social-care records, ensuring more precise, data-driven clinical decisions. Labour-market researchers could track real-time employment and retraining trends, comparing policy interventions' impact on reducing economic inactivity. Climate scientists could integrate air-pollution and health data, pinpointing the direct impact on respiratory illnesses and shaping stronger public-health responses.¹⁵

The NDL would strengthen the UK's position as a global research leader. Russell Group universities generate £38 billion annually, but bureaucratic delays and siloed data limit research impact.¹⁶ Institutions like the Francis Crick Institute, which rely on genomic, clinical and environmental data, could attract more international funding and accelerate breakthroughs in disease prevention.¹⁷ And with growing excitement about the role of AI tools in scientific production and discovery, the NDL can position the UK at the forefront of AI-powered discovery, medical innovation and economic forecasting.

Use Case 2 – Commercial R&D: Revitalising Innovation and Restoring Global Competitiveness

The NDL could turn public-sector data into a catalyst for commercial innovation, driving faster AI development, stronger life-sciences research and a more competitive UK R&D sector. Commercial R&D investment amounts to £50 billion annually in the UK. The NDL would make this investment significantly more productive by giving businesses seamless but safe access to high-quality data, accelerating product development and

global competitiveness. Existing platforms like ADR UK and the IDS have shown the value of linked administrative data, but limited commercial availability has prevented businesses from fully leveraging these data sets.¹⁸ The NDL would reduce these barriers by providing the secure, scalable infrastructure needed to integrate public-sector and third-party data across sectors.

Pharmaceutical researchers could set up clinical trials faster, improving drug development and precision medicine. The NDL could link NHS, social care and genetic data, reducing delays in trial recruitment, regulatory approvals and patient monitoring – potentially reversing the 44 per cent drop in clinical-trial activity since 2017 and adding £250 million annually to the life-sciences sector.¹⁹ In tandem with a National Data Trust for health data, the NDL would provide the broader infrastructure for secure, federated access across domains.²⁰

In AI-driven industries, models could be made more accurate with real-world public-sector data. In education, for example, access to real-world data significantly improves the performance of AI models.²¹ The NDL could link student-progression, employment and earnings records to anonymised assessments and teaching materials, so that developers could train AI tutors and personalise learning platforms, unlocking an estimated £30 billion in net annual benefit in the long run.²²

Use Case 3 – Policy Design: Enabling Evidence-Based Interventions

The NDL provides the missing infrastructure for real-time, data-driven policymaking, ensuring faster responses to economic inactivity, public-health risks and social-mobility challenges. Instead of working with fragmented, outdated data, policymakers would have the tools to detect trends earlier, target interventions more effectively and continuously refine policies based on real-world evidence.

Labour-market interventions could become more precise. The ONS Labour Force Survey's failures have left policymakers without visibility of real-time employment shifts, making it harder to respond quickly to job losses or youth unemployment.²³ The NDL could integrate HMRC, NHS and DWP data,

allowing governments to identify health-related barriers to work and develop targeted employment and welfare programmes that get people back into the workforce faster.

Social policies could become proactive rather than reactive. Household-level data integration could identify financial distress early, enabling targeted support that prevents homelessness rather than reacting to it too late. Education and employment data remain disjointed across agencies, making it difficult to track the long-term impact of skills training and economic policies.²⁴ The NDL would remove these gaps, providing a clearer picture of social mobility so policies target the right barriers, improve access to education and equip people with future-ready skills.

Public-health and environmental risks could be tackled before they escalate. Without integrated data sets, policymakers struggle to link air-pollution exposure with spikes in respiratory illness, delaying action until pressures on the NHS rise. The NDL could change this, allowing decision-makers to track emerging health risks in real time and intervene before hospital admissions surge. Secure synthetic-data integration, generating dummy data that closely mirror the structure of real data sets but replace all personally identifiable information, would ensure that predictive modelling remains possible while protecting privacy, solving a critical barrier in policy areas involving sensitive personal records.

The NDL's impact would be amplified when integrated with the National Policy Twin (NPT), a computational twin that allows policymakers to test interventions before implementation.²⁵ The NDL would provide a clear, recognised framework to access properly linked, high-quality data across departments. Combined with the NPT's computational-twin capabilities, this would enable rapid scenario modelling and iteration, simulating the impacts of policy changes. This could cut planning time from months to days, ensuring that policy decisions are continuously refined based on real-world conditions rather than outdated assumptions.

The NDL would move policymaking from slow, fragmented decision-making to a system that is dynamic, predictive and adaptive. Governments would be able to test, simulate and refine policies in real time, delivering faster, smarter and more effective interventions that improve lives and drive economic growth.

Use Case 4 – Public-Service Delivery: Personalised and Proactive Solutions

The NDL will be key to delivering proactive, personalised public services. As it stands, action is typically taken only after problems escalate, but the NDL could enable earlier interventions, better coordination and more efficient resource allocation.

Local councils would be able to act sooner to prevent poverty and exclusion. By linking social-care, health and welfare data, the NDL could allow councils to identify at-risk families earlier, intervene sooner and prevent crises before they require costly interventions. Without this infrastructure, support arrives too late, increasing long-term costs and worsening outcomes.

Public services would work smarter with integrated data. NHS and employment records would be linked, enabling better-designed re-entry programmes for individuals recovering from health conditions, improving workforce participation. HMRC and employment-data integration could strengthen tax-compliance efforts while reducing administrative burdens on businesses.

Public infrastructure projects could be delivered faster, with lower costs and less disruption. The NDL could enhance the National Underground Asset Register (NUAR)²⁶ by linking planning, construction and utilities data, allowing developers to avoid delays, prevent costly mistakes and improve coordination across agencies.

The NDL would ensure public services are intelligent, coordinated and fit for the future. Instead of reacting to outdated or incomplete data, the government would be able to anticipate needs, target interventions and drive lasting improvements in efficiency and impact.

These four uses cases demonstrate the NDL's potential to unlock data and create a powerful multiplier effect across different ecosystems, where every connection amplifies the next, accelerating breakthroughs and driving smarter decision-making. Pharmaceutical companies could develop treatments faster with linked health and demographic data, while regulators and clinicians could gain real-time insights into effectiveness and patient outcomes. In education, linking pupil assessments with employment outcomes could refine teaching strategies and shape evidence-based policy.

04

The Road to a National Data Library

With such a wide array of use cases and benefits, building the NDL will require striking a fine balance between generating quick wins to build momentum, iterating work based on lessons learned and delivering transformative impact at full scale. At its most expansive, the NDL should underpin a full stack of applications and services capable of reshaping entire public-service value chains. Achieving this vision depends on acting decisively now, not in years to come, while remaining committed to an ambitious agenda.

Immediate Actions (First Six Months): Leveraging Existing Infrastructure

Immediate steps to build the NDL should focus on delivering tangible results by unlocking the value of existing data assets, and concentrate predominantly on anonymised and de-identified data. This does not negate the need for a robust funding case as part of the imminent Spending Review for the development of a fully functional NDL, but DSIT should not wait for the review to be concluded.

Specifically, these steps should address fragmented access regimes and the lack of data quality required for modern AI-driven applications, illustrate the NDL's longer-term potential through quick wins and set the stage for long-term transformation. This initial phase should be a series of sprints linked to specific use cases where the team can learn quickly and fail safely, working with external partners and experts to its capacity.

INITIAL GOVERNANCE AND LEADERSHIP

Building the NDL into a fully operational and trusted institution requires strong early-stage governance and clear leadership structures. The immediate challenge is ensuring effective coordination across government, industry and research stakeholders, while also driving early implementation through a series of high-impact sprints.

In the near term, DSIT should incubate the NDL, providing an initial governance structure and coordinating access to existing data assets. Key to this will be a strong central leadership hub with clear decision-making authority, a dedicated team capable of rapid problem-solving and delivery, and a network of supporters across early-user departments and government bodies.

A senior leadership team must be established with clear ministerial oversight and cross-government coordination. Given the NDL's strategic role in unlocking the power of AI and data for the delivery of the government's priorities, its structure should bring together DSIT, the Treasury and key data-related departments.

Recommendation: Appoint a senior leadership team to oversee NDL implementation. This should include:

- Appointing a chief executive or managing director with sufficient authority to drive change.
- Putting in place a sponsoring Executive Board that would include the DSIT secretary of state, the chief secretary to the Treasury, the national statistician, the prime minister's AI opportunities advisor and representatives from the ESRC, national-security agencies and public-sector users.
- Setting up a Data Access Committee, chaired by a senior official from the UK Statistics Authority, to oversee secure, ethical and scalable data-sharing frameworks.

- Formalising ministerial sponsorship to ensure the NDL has high-level political backing. The prime minister should be regularly briefed on progress towards NDL implementation and any early wins the NDL is generating.

The NDL team itself should consist of:

- Sprint teams to drive early use cases, led by agile delivery specialists to ensure iterative, user-focused development.
- A central service line for users, with a dedicated problem-solving team to unblock early access barriers.
- A user-engagement programme to build trust, incorporating researchers, industry partners and government stakeholders into the design process.

The immediate tasks for the leadership team are to agree on a limited number of high-impact, low-effort use cases, initiate early sprints oriented around them, identify any immediate legislative changes needed to unblock progress and ensure a “whatever-it-takes” approach to removing barriers for potential users. From the start, both the leadership and the delivery team should engage closely with users to correct course quickly.

LEVERAGING EXISTING ASSETS

The NDL is not starting from scratch – the UK already has valuable data assets and platforms that provide a foundation for linked public-sector data. However, existing initiatives remain fragmented, access is slow and commercial use is often restricted. The NDL must build on these resources while ensuring seamless, scalable access and better support for data controllers.

Several key platforms already enable secure data access, including the ONS Secure Research Service (SRS) and the IDS, which provide controlled access to de-identified, linked administrative data sets. Other initiatives, such as the AI Education Content Store, Find Open Data, the Health Data Research Gateway and ADR UK Data Catalogue, improve data-set visibility and research accessibility but lack integration capabilities and broad commercial access. Sector-specific resources such as UK Biobank,

Genomics England and the NHS Research Secure Data Environment Network provide valuable but siloed health-data sets, limiting their usability across research and policy domains.

These resources provide the NDL with usable data from day one, but they are rarely designed for seamless cross-sector integration, preventing wider use of public-sector data to drive economic and social impact. In many cases, data controllers lack the resources and support to prepare data sets for sharing, leading to bottlenecks in data linkage and accessibility.

Recommendation: The NDL should prioritise integrating the ADR UK and IDS data sets, as they are already advanced, linked administrative assets with strong investment and established relationships with data controllers. These data sets can form the foundation for NDL expansion, ensuring an early base of high-quality, linked public-sector data that can be accessed securely and efficiently. The NDL should invest in supporting data controllers to prepare and maintain data sets for broader use rather than focusing investment solely on technical infrastructure.

Recommendation: The government should expand the inventory of pre-linked data sets to include those that have already demonstrated impact, such as the Cross-Justice System data set (Ministry of Justice) and Longitudinal Education Outcomes (LEO) data set.^{27,28} These data sets, when integrated into the NDL, will provide richer insights into justice-system patterns, educational outcomes and economic mobility. Alongside this, bespoke linkage services should be developed to meet evolving research and industry needs, ensuring that the NDL remains adaptable and responsive as new data requirements emerge.

EMBEDDING EXPERTISE TO DRIVE DATA READINESS AND USABILITY

The NDL must act quickly to unlock data, improve usability and ensure public-sector data drive real-world impact. Current challenges exist on both the supply and demand sides. Departments struggle to make data available efficiently, while researchers and businesses lack the support needed to access and apply administrative data to high-impact questions.

A key barrier is the lack of cross-government coordination in data preparation and sharing. Without dedicated expertise embedded within departments, the NDL risks being underutilised and high-value data sets may remain locked away in bureaucratic silos. Departments need direct, embedded support to improve data quality, enhance usability and align data sets with the needs of researchers, policymakers and industry.

Recommendation: The NDL should create a network of National Data Librarians embedded within key government departments and services, including the NHS. These officials will act as cross-government advisors, ensuring that data are accessible, usable and aligned with real-world needs. Modelled on the chief scientific advisers (CSAs) network, National Data Librarians will bridge the gap between government data controllers and external users, ensuring the NDL delivers practical value from the outset and that each relevant government body has a person responsible for making its data safely accessible through the NDL.

While the National Data Librarian role draws inspiration from traditional librarianship principles, it represents a distinct governmental function not governed by existing library accreditation frameworks. This role will not be defined by conventional library science credentials and should be exempt from any associated regulation.

A successful National Data Librarian network must balance institutional continuity with external expertise, ensuring the NDL remains adaptable while embedding long-term capacity within government. In the early stages, the NDL would benefit from an injection of fresh perspectives and technical expertise from outside government. Over time, the network should transition into a stable, permanent structure that retains knowledge and builds sustainable departmental expertise.

Recommendation: In the initial phase, National Data Librarians should be brought in on three-year secondments from academia or industry, ensuring a continual influx of expertise and fresh thinking. This will also help to extend the NDL's external network, ensuring deeper engagement across government, research and commercial sectors.

Recommendation: To ensure institutional continuity, a core team of permanent National Data Librarians should be established from year three onwards, providing departmental expertise and stable leadership. The transition should be structured so that some secondees shift into permanent positions, while others are replaced by new external talent to address the NDL's talent needs as these change over time.

For the NDL to be effective, National Data Librarians must operate within a structured, accountable framework. Their role will include identifying opportunities for data linkage within and across public bodies, ensuring key data sets are connected to maximise impact. Within departments, they will advise on how to leverage the NDL for policymaking and service delivery, embedding data at the heart of decision-making rather than leaving them underutilised. They will also coordinate with UK Research and Innovation (UKRI) to identify research-ready data sets and inform targeted funding calls.

National Data Librarians will also ensure cross-departmental coordination around data usability and provide support for data controllers in structuring data sets for broader use. Beyond government, they will engage with third parties, acting as a bridge between government, academia and industry to ensure data sets meet user needs. Their role will also include horizon scanning for innovation opportunities, ensuring the NDL remains adaptable to emerging data challenges and policy demands.

Without clear leadership and coordination, this wide range of activities risks becoming disconnected from real-world needs. A Chief National Data Librarian is needed to lead this network, ensuring that National Data Librarians are strategically positioned, aligned with government priorities and equipped to drive impact across sectors.

Recommendation: To align NDL strategy with government priorities, National Data Librarians from relevant departments and bodies should be invited to join mission boards to address data gaps and ensure linkage opportunities feed into mission delivery.

Recommendation: The network should be led by a Chief National Data Librarian, working closely with the government's chief scientific advisor, the national technology advisor and the No. 10 AI advisor to keep the NDL aligned with broader government priorities on AI, data and digital transformation. As the NDL transitions into an arm's-length body (ALB), a dedicated permanent role in No. 10 should provide strategic oversight, ensuring long-term accountability and impact.

To scale the NDL, departments need both incentives and support to share data faster. Embedding National Data Librarians within early adopters will ensure that those committed to data sharing receive direct expertise and resources, improving data quality and accessibility. At the same time, funding must be rebalanced to support data controllers, ensuring that investment goes beyond infrastructure to the people responsible for preparing and linking data sets.

Recommendation: To incentivise participation and drive early adoption, National Data Librarians should first be embedded within departments that commit to faster data sharing, ensuring that those proactively engaging with the NDL benefit directly from additional expertise and resources. Funding should be set at £10 million per year, covering 50 roles across government and key public-sector bodies, with an average cost of £150,000 per role (including pensions and National Insurance contributions).

This could be funded by rebalancing core ADR and IDS budgets, directing more resources towards data controllers responsible for preparing and linking data sets. To make this viable, the benefits to ADR and IDS must be clearly demonstrated, showing how better-prepared data sets will reduce inefficiencies, speed up approvals and improve research output. This approach shifts ADR and IDS from gatekeeping to enabling, strengthening their role in an efficient, well-governed data ecosystem.

REMOVING BARRIERS TO DATA SHARING AND ACCELERATING ACCESS

Even with the right expertise and infrastructure in place, departments may remain hesitant to share data due to legal complexities and potential liabilities. Without clear protections, risk aversion will continue to slow down

access to data. Addressing this challenge requires a formal indemnity framework to give data controllers the confidence to share responsibly within established legal safeguards. A precedent for this exists: during Covid-19, emergency measures allowed general-practice data to be shared more comprehensively, supported by legal safeguards and government-backed indemnities. A similar approach could help departments overcome legal and reputational concerns, particularly in complex areas such as health, tax and justice data.

Recommendation: The government – through a secretary of state, the prime minister or the Cabinet Office – should establish an indemnity scheme for data controllers. This would reassure departments that the risks of sharing data for linkage under the IDS and ADR schemes, and eventually via the NDL, are low, reducing hesitancy and ensuring more consistent data availability. The scheme should be insurance based, providing financial and legal protection where necessary, while also driving convergence towards more common data standards across departments.

Even with stronger legal protections and governance, the process of requesting and accessing public-sector data remains slow and fragmented. Multiple departmental data-access committees create duplication, inconsistency and unnecessary delays, discouraging users. Evidence shows that cutting lead times from months to weeks significantly increases demand, as seen with London's Discover-NOW trusted research environment (TRE), which saw greater uptake after reducing delays.²⁹ A single, streamlined approval process following the established "Five Safes" approach³⁰ would ensure that data are made available faster while maintaining accountability and security.

Recommendation: To increase demand and reduce transaction costs, the NDL should establish a single, centralised Data Access Committee (DAC) with delegated authority from departmental committees for all data within the scope of the NDL. This would eliminate duplication, streamline approvals and ensure faster, more consistent decisions while maintaining accountability.

The DAC should be set a goal of radically speeding up access to data, so that within the first 24 months, all data-access requests should be processed within two weeks of submission. The DAC should adopt a risk-stratification approach that segments data-access requests according to the track record of the requester (for example, a civil servant already vetted within government versus a commercial user outside government) and the class of data and/or analysis requested (for example, personal data versus non-personal data or a routine versus novel research question).

To ensure rapid adoption and build momentum, the NDL must deliver tangible benefits from the outset. Rather than waiting for full-scale implementation, a targeted approach focusing on a small number of high-impact use cases will allow the NDL to refine its processes, address operational challenges, and showcase its value across government, research and industry. These early projects should be aligned with the five government missions, ensuring that data-integration efforts directly support national priorities.

Recommendation: A small number of high-impact use cases, linked to the five government missions, should be identified for “minimum viable products” to be developed and delivered via agile sprints to address the specific operational challenges of creating and using linked administrative data.

LAYING THE LEGAL GROUNDWORK FOR SECURE AND SCALABLE DATA SHARING

Ensuring the NDL has a strong legal foundation is critical to enabling effective data linkage while maintaining public trust. A clearer legal framework is required to expand data use beyond its current limitations. The ONS’s legal mandate under the Statistics and Registration Service Act 2007 provides a strong foundation for data linking but is currently limited to aggregated statistical analysis for policy and research.³¹ The current Data (Use and Access) Bill³² (DUAB) provides an early opportunity to introduce targeted legislative changes that can facilitate data integration before a more comprehensive National Data Library Bill is introduced in the medium to long term.

Health data present a specific legal and operational challenge. Given the complex regulatory landscape, we do not anticipate health data being available through the NDL in its first phase. In a [previous report](#), TBI proposed a separate National Data Trust for health data, and the recent Sudlow Review has recommended the establishment of a National Health Data Service for health, aligning with this approach.³³ However, the NDL should still lay the groundwork for future integration by addressing key legal barriers now.

As the Sudlow Review makes clear, there is a reluctance among many health-care professionals to share data due to “concerns about the time needed to engage with data sharing and the potential legal liability (for example for inadvertently breaching data-protection laws or obligations under the common law duty of confidentiality)”. This is further complicated by the complex interplay between the common-law duty of confidentiality and UK data-protection law.

Complying with the common-law duty of confidentiality can be challenging in circumstances where it is not possible or practicable to obtain consent from individuals. While regulations allow for certain uses of data without consent, the scope and interpretation vary considerably across England, Wales, Scotland and Northern Ireland, leading to a fragmented legal landscape. A broader, more unified, cohesive approach to addressing the common-law duty across the UK where patient consent is not feasible would have clear benefits for research using health data.

The DUAB process offers a window to make administrative- and health-data linkage more straightforward, even if direct integration is not yet feasible. To move forward, two main legal avenues could be pursued:

Recommendation: The government should consider the relative merits of two legislative options to enable more seamless health-data linkage.

- Amend section 65(4) of the Digital Economy Act 2017 (DEA) to include health and social-care data. The DEA provides a gateway for the voluntary disclosure of information held by a public authority to another person for the purposes of accredited research but currently it does not

apply to public authorities with functions relating to the provision of health services or adult social care. Expanding the DEA's scope could simplify linking health data sets with other administrative data. However, while the DEA provides a voluntary legal basis for sharing, it does not compel data sharing and therefore has inherent limitations particularly where there is a cultural reluctance to share data.

- Use the Health Service (Control of Patient Information) Regulations 2002 (COPI), which were employed during Covid-19 to enable rapid data sharing.³⁴ This would allow the secretary of state to issue a notice to data controllers, requiring them to share relevant health data. This mechanism is stronger than the DEA amendment but would require careful oversight.

Ensuring public confidence in data sharing is just as important as establishing a strong legal foundation. Past changes to health-data legislation have triggered waves of opt-outs, making the data sets less representative and limiting their usefulness for research and policy. At the same time, legal uncertainty over what constitutes confidential data creates hesitation among data controllers, further slowing progress. There is an urgent need to clarify aspects of the UK data-protection law and in particular definitions around anonymisation. Addressing these issues is critical to ensuring that health data remain both accessible and trustworthy.

Recommendation: The government should revise the health-data opt-out process to make it significantly easier for individuals to opt back in. Patients should be given more information about why it is important that data reflect the diversity of the whole population to encourage them to opt back in.

Recommendation: The Information Commissioner's Office (ICO) should issue clear, detailed and definitive guidance (including case studies) on what constitutes anonymous data under data-protection laws, specifically covering scenarios where data are held and processed within a TRE with direct identifiers removed, as well as synthetic data sets generated from originally identifiable data. In drawing up this guidance, the ICO should be led by the government's clear steer on the need to "regulate for growth".

DEMONSTRATING TRUSTWORTHINESS

The NDL must actively build and maintain public trust. Without strong transparency and engagement, the NDL risks being seen as intrusive or misaligned with public expectations, undermining its long-term viability. Trustworthiness requires clear visibility into how data are used, and meaningful public participation in shaping governance and policy decisions.

Public confidence in the NDL will rely on clear, accessible information about how data are being used, by whom and for what purpose. As TBI has [previously set out](#), evidence suggests that the public is broadly comfortable with data sharing when it is secure and demonstrably beneficial, but this level of transparency is currently difficult to obtain. Without proactive disclosure, misinformation and distrust could take hold, jeopardising engagement and adoption.

Recommendation: The NDL should establish a comprehensive transparency layer from the outset by creating a public registry of approved projects, data sets accessed and their intended use. Initially, this can follow the approach taken by the United States's National Artificial Intelligence Research Resource, publishing and regularly updating a spreadsheet with a list of projects.³⁵ Over time, this should evolve into an interactive, queryable platform that allows the public to track how data are applied and the benefits delivered.

Trust in data sharing is not a one-time achievement but must be continuously reinforced through transparency, safeguards and citizens' participation. Past failures, such as the care.data initiative to pull together patient data from GP practices, highlight the risks of overlooking public concerns, while successful initiatives such as NHS England's Data for Research and Development programme and Estonia's X-Road platform demonstrate that early and meaningful engagement fosters trust.³⁶ The NDL must go beyond traditional consultation, ensuring that citizens play an active role in shaping key policy and design decisions.

Recommendation: The NDL should embed public engagement into its core policy-design teams, ensuring that societal values and concerns are reflected in decision-making from the outset. It should adopt proven co-design frameworks, such as those used in OneLondon, to structure engagement processes.³⁷ These frameworks should focus on privacy mechanisms, access rules and the commercial model, aligning the NDL with public expectations and ethical standards.

Recommendation: The government should allocate dedicated resources and time to citizen-participation efforts, ensuring engagement is meaningful and well supported. This should include collating evidence from previous public deliberations on data sharing, designing robust, evidence-based engagement initiatives, and prioritising outreach to underrepresented groups to ensure diverse perspectives are included.

Medium-Term Actions (Six Months to Three Years): Expanding the NDL's Capabilities

The second phase of the NDL should focus on widening access, improving technical infrastructure and capabilities, building a commercial strategy for access and preparing an NDL Bill to provide the necessary legal basis for the organisation. This would include improving the user experience and widening the service offer, as well as expanding its use cases beyond academic research and intra-government applications.

There are several components to the NDL service offer. These will require procurement, and there are different approaches the government might want to explore, some already set out in the Wellcome Trust and ESRC white papers discussed earlier in this report.

ESTABLISHING A SECURE AND SCALABLE ACCESS FRAMEWORK

A well-structured access system will ensure trusted users engage with public-sector data securely and efficiently, enabling researchers, policymakers and industry to maximise value. A clear, flexible permissions framework will maintain oversight and accountability while streamlining approvals, reducing delays and driving adoption.

The NDL must balance security with usability through a tiered access system, granting appropriate permissions based on credentials, compliance and prior usage. At the moment, users can spend months securing access to very similar types of data sets within the same or closely related institutions. The NDL should introduce a dynamic and adaptable system, ensuring seamless access without unnecessary bureaucracy, while maintaining strong governance to protect sensitive information. This is a key requirement for its longer-term success.

Recommendation: The NDL should introduce a Reader Pass (a kind of digital-library membership card), defining user permissions based on credentials, compliance and past usage. This would involve several tiers of access, ranging from a basic user account streamlining open-data access to researchers who have completed information-governance training or civil servants operating under safeguards such as the Official Secrets Acts.

The Reader Pass should define which groups of data sets users can access under which conditions, implementing a passporting approach under which permissions can be shared across different NDL assets at the same tier, subject to continued compliance. This would ensure that data are used efficiently and accessed quickly while remaining secure. Access rights should be monitored and, if necessary, adjusted over time so that security measures align with users' risk profiles and the nature of their work. The Motivated Intruder Test framework should be used for risk assessment of data sets, ensuring that safeguards are proportionate to the sensitivity of the data being accessed.³⁸

A secure access framework requires clear accountability to maintain trust and drive adoption. Enforceable terms of use and oversight mechanisms will prevent misuse, while streamlined processes will give data controllers the confidence to share information. Overly restrictive policies slow innovation, but a balanced approach will ensure security without creating unnecessary barriers to access. At the moment, much of the risk of data misuse sits with the data controllers while the benefits of use accrue to those accessing the data. Disincentives for misuse by the latter can be introduced alongside the Reader Pass. A secure access framework will also reduce the risk aversion of data controllers, providing clear visibility over who is accessing which data and for what purpose.

Recommendation: To maintain accountability, the NDL should mandate users to commit to clear terms of use, with enforceable consequences for misuse. These should include suspension or permanent revocation of the Reader Pass for unauthorised data sharing, and the introduction a Data Offenders Register for those who abuse this trust. A transparent review and appeals process should be in place to ensure fairness and maintain trust. Trusted users should benefit from passported access, eliminating the need for repeated approvals for similar data requests. This approach will streamline access, reduce administrative delays and foster a culture of responsible data use, ensuring that security does not become a barrier to innovation and policy.

BUILDING A METADATA CATALOGUE FOR SEAMLESS DATA DISCOVERY ACROSS THE PUBLIC SECTOR

A comprehensive metadata catalogue is essential to making the NDL accessible and user-friendly. Without a clear, searchable index, data sets risk being underutilised, difficult to navigate or inaccessible to those without technical expertise. A well-designed metadata system will allow users to quickly locate data sets, understand their content and sensitivity, and access them efficiently.

The NDL's metadata catalogue will function as the library's index for public-sector data, providing essential information about each data set, including its contents, permitted uses and security classification. Unlike existing

platforms, which often require multiple applications across different custodians, the NDL will consolidate discovery and access into a single, intuitive system. The experience of London's Datastore programme shows that using a basic shared metadata schema helps not only to simplify discovery but also to ensure the included data sets use a similar structure over time.³⁹

Recommendation: The NDL should establish a metadata catalogue that allows users to efficiently discover and understand available data sets, functioning like a library index. The Incubator for Artificial Intelligence, which sits within DSIT, should develop a search interface that would allow users with a Reader Pass to search for data they need using plain language (similar to its Lex pilot), making the system accessible even to those without technical expertise.⁴⁰ The catalogue should be maintained and serviced by the network of National Data Librarians, ensuring that data descriptions remain accurate and machine-readable, access rules are clear and usability improves over time. Technology should be leveraged to automate routine processes, allowing for scalability and efficiency.

By centralising data-set discovery, the NDL will remove inefficiencies that currently require users to navigate multiple custodians and approval processes. A researcher looking for educational data, for example, should be able to quickly locate anonymised pupil-assessment data sets, view clear access rules based on their own rights and understand how they can be used. Existing platforms, such as the National Pupil Database⁴¹ and the Health Data Research Gateway,⁴² provide some of this functionality but remain fragmented. The NDL should unify and simplify this process, ensuring that public-sector data are as easy to navigate and use as possible.

The NDL should also carry out an audit of existing data sets where access limitations put a brake on research and innovation within the private and public sectors, and support their discovery by users. (A good example of a welcome step in this direction is the publication of a basic catalogue on Github by the Incubator for Artificial Intelligence).⁴³ It should also engage early on with existing projects around non-administrative data in government, such as the team behind the NUAR and the National Digital Twin Programme, to ensure that they are discoverable through the NDL.

Finally, the NDL should develop business cases for making strategic data sets more easily accessible or free to use. The Postcode Address File could be an early example of this, with similar open-data initiatives for address data in Denmark delivering a 1:31 cost-benefit ratio.⁴⁴ The NDL would be a natural home to house and maintain data sets of this nature.⁴⁵

Recommendation: The NDL should over time expand its remit to include non-administrative (for example, geospatial) data, improving discovery for users and supporting open-data initiatives for strategic data sets that would support economic growth and innovation.

PROVIDING ANALYTICAL SUPPORT TO MAXIMISE DATA USABILITY

Access to data alone is not enough – users need the tools and support to use these data efficiently. Without integrated analytical capabilities, researchers, policymakers and businesses may struggle to process and interpret complex data sets, limiting the NDL's impact. A comprehensive analytical service will ensure that users can navigate, analyse and apply public-sector data effectively, reducing technical barriers and enhancing usability.

The NDL's analytical support should go beyond data discovery, offering automated tools and expert assistance to help users prepare, process and interpret data. This will ensure that data sets are not only accessible but also actionable, making public-sector data more valuable for decision-making and innovation.

Recommendation: The NDL should develop a comprehensive analytical service that enables users to navigate and process data sets efficiently. A dedicated service desk should provide automated support, handling tasks like data wrangling and basic analysis through advanced digital infrastructure, AI tools and visualisation tools. Data should be available in standard formats to be analysed with common programming languages such as Python. This will remove technical barriers, allowing users to focus on insights rather than time-consuming data preparation.

Recommendation: The National Data Librarian network should play a key role in enabling collaboration with external technology partners, ensuring that the NDL remains at the cutting edge of analytical capabilities. This will help further enhance the usability and value of data sets, ensuring that users can extract meaningful insights without requiring deep technical expertise.

Recommendation: The NDL's analytics layer should be designed as a modular system, allowing for flexibility in how users interact with data. Whether through a single integrated platform or a suite of analytical products, this system should be informed by user needs and international best practices, ensuring it remains adaptable and scalable over time.

BUILDING A SCALABLE AND FUTURE-PROOF DATA ARCHITECTURE

To ensure long-term success and adaptability, the NDL must be built on a flexible, modular architecture that supports secure data linking, governance and streamlined access. A rigid, centralised model risks repeating the failures of past large-scale government IT projects, whereas a federated and modular approach allows for incremental improvements and integration with emerging technologies.

Recent work by the ESRC and the Wellcome Trust has highlighted multiple viable architectural options, with broad consensus that the NDL should prioritise a federated structure. However, this approach must be balanced with the need to ensure machine-learning (ML) compatibility, as federated systems often pose challenges for ML applications.⁴⁶ This can be mitigated by providing pre-linked data sets within the NDL or through architectural design choices that facilitate AI-readiness. The Wellcome Trust Technical White Paper Challenge process has outlined possible solutions to navigate this trade-off while maintaining AI usability.

Recommendation: The NDL should adopt a minimum-viable-product (MVP) approach, prioritising foundational capabilities such as secure data linking, governance frameworks and streamlined access controls. Additional functionality should be developed iteratively based on real-world use and user needs. The NDL should optimise its federated approach for AI applications, ensuring that data remain usable for machine learning while

upholding privacy and security standards. It should integrate pre-linked data sets where necessary and explore architectural innovations that enhance AI readiness while maintaining decentralised control over sensitive data.

As the NDL evolves, synthetic data will play a critical role in expanding access while ensuring privacy compliance. By generating AI-ready data sets that preserve statistical accuracy without exposing personal records, synthetic data will enable advanced analytics while adhering to strict governance and privacy standards.

However, synthetic data are not without risks – research has shown that in some cases they remain vulnerable to reconstruction or inference attacks if not designed carefully. The degree of risk depends on the generation methods used and the level of statistical fidelity required.⁴⁷ To overcome these challenges, the UK must invest in next-generation approaches to privacy-preserving synthetic data.

Recommendation: The NDL should collaborate with the Advanced Research and Invention Agency (ARIA), including through match-funded work, to develop advanced synthetic-data-generation techniques, ensuring that privacy is maintained while preserving the utility of data for AI applications. ARIA's mission to fund high-risk, high-reward research makes it an ideal partner for tackling the complex challenges of balancing privacy, accuracy and scalability in synthetic data. ARIA's leadership should consider incorporating this work as a new opportunity space within its Safeguarded AI programme.

Recommendation: UKRI, in collaboration with the NDL and relevant government bodies, should leverage existing research on disclosure-control measures within TREs to enhance synthetic-data safeguards, ensuring alignment with best practices in AI governance and privacy protection.⁴⁸

CREATING HIGH-VALUE DATA SETS FOR SCIENTIFIC AND INDUSTRIAL INNOVATION

To drive long-term value for academic and industry users, the NDL must go beyond linking existing administrative data sets and actively support the creation of new, high-value data sets that meet the needs of researchers, businesses and policymakers. As TBI has [previously described](#), without a coordinated effort to develop data sets that do not currently exist, the UK risks missing opportunities to drive breakthroughs in science, industry and public services.

Public-private collaborations, such as Our Future Health and UK Biobank, already position the UK as a leader in data-driven life sciences. However, critical gaps remain in scientific and industrial data sets, limiting the potential for new discoveries and applications in areas such as drug discovery, AI-powered materials science and environmental monitoring. The NDL should coordinate efforts to identify and fill these gaps, ensuring the UK remains at the forefront of data-driven innovation.

Recommendation: The NDL should work with UKRI and the technology industry to orchestrate targeted data-creation challenges in areas of global importance, similar to the Critical Assessment of protein Structure Prediction (CASP) protein-folding challenges.⁴⁹ These challenges should identify and fund the development of high-potential data sets that currently do not exist but could unlock breakthroughs in science and technology. Collaborations between academia and industry should be a core part of this initiative, ensuring that data sets are designed with end-user needs in mind.

In life sciences, for example, new data sets could include comprehensive databases of peptide sequences to accelerate drug discovery or advanced cellular atlases to transform regenerative medicine. The NDL should facilitate the systematic collection and integration of cellular-level data, enabling breakthroughs in disease treatment and human biology.

For particularly ambitious projects, the NDL should partner with ARIA, including through match-funding efforts, to build high-impact data sets that require collaboration across multiple institutions. ARIA's mandate to fund

transformative research – exemplified by the Scoping Our Planet opportunity space, which focuses on new approaches to monitoring the planet⁵⁰ – makes it an ideal partner for supporting complex, cross-disciplinary data initiatives.

The next frontier in data-driven innovation lies in creating, deploying and integrating sophisticated sensor networks. Advances in quantum sensors, biosensors and other emerging technologies are enabling entirely new ways of monitoring the environment, tracking molecular interactions and detecting minute changes in real-world conditions.⁵¹ However, data from these diverse sources must be properly structured, integrated and made actionable – a role the NDL is uniquely positioned to support.

Recommendation: The NDL should develop frameworks for integrating next-generation sensor data, ensuring these emerging data sets are structured for interoperability and usability. This effort should align with ARIA's Scoping Our Planet initiative, which is already funding new approaches to Earth-system measurement and high-resolution data collection. By integrating AI-ready sensor data into the NDL, the UK can develop a fundamentally new way of understanding and responding to scientific, industrial and environmental challenges.

This capability will become particularly powerful when combined with AI systems capable of processing multiple sensory streams simultaneously, revealing patterns and relationships that remain invisible to traditional analysis methods.

BUILDING SUFFICIENT COMPUTE CAPACITY FOR AI-DRIVEN RESEARCH

The NDL's ability to support AI-driven research and large-scale data analysis will depend on the computing power, usually referred to as "compute", available to its users, including through a dedicated trusted research environment. Without adequate compute infrastructure, the full potential of linked public-sector data for ML and advanced analytics may remain unrealised. Making data AI-ready requires vectorisation – converting information into numerical representations that machines can process. This

transformation enables AI models to analyse and learn from diverse data types efficiently. But this process is compute-intensive, especially for large data sets, such as the one-terabyte LEO data set.

The forthcoming DSIT compute strategy, set for publication in the spring of 2025, presents an opportunity to align the NDL's infrastructure needs with the UK's long-term approach to sovereign compute. Existing government efforts around compute capacity that is either owned outright by the public sector or allocated by it should be optimised as a shared asset for the NDL. This would drive down costs, avoid vendor lock-in and reduce fragmented procurement across departments. Leveraging academic on-premises compute capacity as part of a hybrid cloud/on-premises model would further strengthen scalability.

The AI Opportunities Action Plan recommends coupling compute allocation with access to proprietary data sets. While neither the plan nor the government response to it provide further implementation details, one approach could be to reserve part of the AI Research Resource (AIRR) – the UK's national AI supercomputing infrastructure, which is set to expand 20-fold as recommended in the plan – for NDL users.⁵²

AI projects often face critical bottlenecks when compute needs outpace initial resource allocation. To prevent promising innovations from stalling at the prototype phase, the NDL's compute allocation to projects should be dynamically adjusted based on project milestones and outcomes.

Recommendation: DSIT should explicitly factor the NDL's requirements into its forthcoming long-term compute strategy, ensuring that AI-driven research within the NDL is not constrained by compute limitations. A significant fraction of AIRR's compute capacity should be reserved for NDL users, with flexible scaling mechanisms based on project performance. This could include automated resource allocation, with projects showing early success qualifying for increased compute capacity – for example, a threefold increase in allocation for projects demonstrating strong initial results. The Reader Pass can act as a credential to prove that a user is entitled to use their allocated share of AIRR compute.

DEVELOPING A SUSTAINABLE COMMERCIAL STRATEGY

As third-party private-sector access to the NDL expands, a clear and sustainable commercial model must be established. The NDL must balance financial viability with its role as a public good, ensuring that access remains broad and equitable while generating sufficient revenue to cover the cost of operations. The model should be designed to reflect the diverse needs of users, from government and research institutions to private companies, both domestic and international. The primary aim of the NDL's commercial model should be to recoup development and operational costs, not to generate profit. This aligns with its role as vital enabling infrastructure.

Revenue generation must be proportionate to the value derived, ensuring that government use remains free while commercial users contribute in a way that supports the NDL's long-term sustainability without prioritising profit maximisation. This will require a carefully structured fee system, informed by public engagement and stakeholder consultation.

Recommendation: The NDL should introduce a structured access model that reflects the varying needs and financial capacities of its users. Government use and access to data sets designated as open data should remain free, while private-sector users contribute based on the value they derive. The Reader Pass for non-government users could include a nominal administrative fee, supporting the infrastructure needed to manage access while keeping the model simple and accessible (though the fee could be waived in some cases, such as for recipients of Global Talent visas).

For large-scale users, a tiered data-access fee structure should be developed, based on the type of organisation or individual accessing the data. Large private companies should contribute more, reflecting the commercial value they gain from enhanced insights and AI-model training, while academic institutions, small organisations and independent researchers should benefit from discounted rates. Institutional subscriptions should be introduced, allowing organisations to provide access to affiliated researchers and employees without direct individual costs, similar to how universities provide affiliated researchers with free access to academic

journals. Such a model is strongly supported by previous public engagement and is already in place in some TREs. The Scottish government also has a tiered pricing model.⁵³

The commercial model must also account for the value of advanced analytical services. The ability to tap into the AIRR's computing resources for AI-model training and large-scale data analysis within the NDL's secure environment is a significant asset beyond data access itself. To ensure financial sustainability while maintaining broad accessibility, the NDL should capture a fraction of the long-term commercial value these assets generate upfront, ensuring that both access and analytics contribute to cost recovery.

The NDL will position the UK as a leader in the global AI and data economy, strengthening its competitiveness in the international AI race. While access for domestic users must be carefully structured, consideration must also be given to international users in academia and the private sector, ensuring that the UK benefits from the global use of its data resources.

Recommendation: If NDL data significantly contribute to intellectual property licensed to non-UK entities, a royalty-payment stream could be introduced, modelled on ARIA's overseas funding mechanism,⁵⁴ ensuring that the UK retains a stake in the economic value of its public-sector data. However, the costs of administering and monitoring such schemes must be carefully evaluated, ensuring that the benefits outweigh any administrative burden. Additionally, NDL access should be positioned as a strategic incentive for attracting top-tier global talent. Providing high-value data sets and advanced analytical tools as part of the Global Talent visa scheme would make the UK a more attractive destination for leading researchers and entrepreneurs, including AI specialists, reinforcing its position as a global hub for AI and data-driven innovation.

INTRODUCING AN NDL BILL TO PUT THE LIBRARY ON A SECURE LEGISLATIVE FOOTING

The NDL's status as a manifesto commitment by the current government signals strong political will, but it also creates risk. Without firm legal foundations and demonstrated value, a change in government could cause the NDL to be deprioritised or even cancelled, limiting the UK's ability to harness AI-linked economic and public-service transformation.

The NDL is vital enabling infrastructure foundational to future progress in AI, research and innovation. To ensure long-term certainty, stability and cross-government adoption, it must be protected from political cycles and ad-hoc decision-making. A clear statutory basis will guarantee its role as a long-term public asset, ensuring sustained funding and operational independence.

Recommendation: The government should put the NDL on a statutory footing and transition it into an arm's-length body (ALB), sponsored by DSIT, through a new Act of Parliament. After the initial incubation period, DSIT, in coordination with other departments and relevant public bodies, should lead the drafting of a comprehensive NDL bill. This legislation should address outstanding legal and governance issues, including:

- A clear statement of the NDL's purpose and objectives.
- Opt-in/opt-out mechanisms for data use to ensure public trust and transparency, integrating lessons from engagement over the early years of its operation.
- Legal provisions for the use of identifiable data and unique identifiers, where necessary, aligning with existing legislation such as the Data (Use and Access) Bill.
- The formal legal structure of the NDL, ensuring its role as cross-government infrastructure rather than being siloed within a single department.
- The legal framework for the commercial model behind the NDL.

- Continuous senior sponsorship and external input into the NDL work and strategic plans, establishing a permanent board that includes secretaries of state of DSIT and other relevant departments, the national statistician, and representatives of UKRI, the AI Security Institute, industry and civil society (such as the Open Data Institute).

The NDL Bill should be passed before the next general election to secure its place in the UK's long-term data and AI strategy.

By design, the NDL will be a cross-government and nationwide resource, meaning it should not sit under a single core department. While transitioning into an ALB ensures greater autonomy, some may be concerned that it could lose its embeddedness within government. This risk should be mitigated by the existing network of National Data Librarians, who will have already been firmly established across departments before the NDL attains ALB status, and continued senior sponsorship of the NDL by the prime minister, DSIT, the Treasury and key data-related departments.

Recommendation: As an ALB, the NDL will gain operational independence, but DSIT ministers should remain accountable to Parliament for its use of public money and performance. This ensures a balance between autonomy and oversight, allowing the NDL to function as a strategic national asset while retaining strong cross-government engagement.

Recommendation: The NDL Bill should define the specific qualifications, expertise requirements and governance framework for the National Data Librarian to reflect. It should codify lessons learned over the first few years of the NDL incubation within DSIT and avoid defaulting to existing library-accreditation frameworks or professional-body standards.

Long-Term Actions (Three to Five Years): Full NDL Implementation

The final phase of the NDL development will see the library become an independent ALB in law following the introduction of the NDL bill. The NDL should continue to be sponsored by the prime minister given the pivotal role

data play in enabling the government's missions and industrial strategy, with regular reports provided on its progress and impact. Important design decisions will already have been made at this stage. There are several international examples that should be used to inform these decisions.

Estonia's X-Road platform is a distributed system that facilitates secure data exchange between government systems and external parties, acting like a digital postal service. It ensures interoperability, strong security and citizens' control over data access. However, it also highlights a key challenge: while it enables secure sharing, it still requires lengthy paperwork and compliance protocols between data custodians and recipients.

Similarly, New Zealand's Integrated Data Infrastructure (IDI)⁵⁵ has demonstrated how a centralised but secure data platform can drive evidence-based policymaking and research. By linking de-identified data sets across government agencies, including administrative, survey and census data, IDI has unlocked major insights in health, education and social policy. It has also encouraged greater uptake of long-term thinking about the outcomes of policy decisions, shifting the mindset of the Treasury towards assessing social rather than purely fiscal outcomes. However, its primary focus remains government use and academic research, with limited accessibility for industry.

The NDL builds on the strengths of both models while going further. Unlike X-Road, it will be embedded within government but designed for broader accessibility – not just facilitating data exchange but accelerating its responsible use across research, policymaking and commercial innovation. Unlike the IDI, it will be structured to enable external access to AI-ready, high-value data sets while maintaining robust governance and security standards.

EXPANDING THE NDL TO INCLUDE IDENTIFIABLE DATA

As the NDL matures, it must evolve to support the use of identifiable data for high-impact applications, where anonymised data sets alone are insufficient. Certain government-service delivery functions, policy-formulation efforts and targeted research projects require a more granular level of detail to be

effective. In particular, there is growing demand among local government authorities to access local residents' data, currently locked away in Whitehall, in order to provide better preventative and proactive services. However, integrating identifiable data within the NDL's more streamlined data-access processes introduces heightened privacy and security considerations, requiring strict governance, clear safeguards and a well-defined access framework.

The NDL cannot function as a repository for identifiable data. Instead, it should with time enable secure, federated linking of data across government departments, ensuring that data remain at their source while being linked on demand. For cases where real-time processing is essential – such as machine-learning applications for public-service delivery – the NDL should support pre-linked data sets for specific service-delivery functions.

The use of identifiable data requires a structured, tightly controlled governance framework. Access should be limited to accredited users, including government officials involved in policy formulation and service delivery, as well as approved researchers conducting legally sanctioned studies. Strong safeguards must be in place to protect privacy while allowing responsible data use.

Recommendation: The NDL should establish clear governance and security measures for handling identifiable data, ensuring that access remains tightly controlled while enabling responsible use. Strict access controls should be in place to ensure that only accredited officials and approved researchers can access identifiable data sets via the NDL, limiting exposure to those with a legitimate policy or research purpose. Data-sharing agreements between data controllers must define specific rules on how data can be accessed, used and stored, creating a legally robust framework for secure data linkage.

All access requests should undergo rigorous evaluation by the NDL's Data Access Committee, ensuring they meet legal, ethical and security standards before approval. Where necessary, the NDL should offer cohort-identification services, allowing for screening and participant identification for health and employment research, but only in cases where this is legally required and based on explicit consent.

The NDL's value in handling identifiable data will not be in centralising access, but in serving as a trusted centre of excellence for linking and analysing data securely. Departments will continue to use linked data within their own operational platforms, ensuring that service delivery remains efficient and aligned with their mandates.

Recommendation: The NDL should act as a central service provider for public-sector applications, ensuring that identifiable data are linked securely and efficiently across government. It should support departments in reducing duplication of effort, allowing them to access linked data without unnecessary administrative burdens. Strong compliance with security and privacy standards must be maintained, ensuring that individual data rights are protected while enabling responsible use. It should provide a structured environment where identifiable data are linked but never leave their secure infrastructure, reinforcing transparency, accountability and public trust in the system.

A UNIQUE PERSONAL IDENTIFIER

Effective data integration across central and local government depends on harmonised personal identifiers. The UK currently relies on disparate systems, such as the NHS number and the national insurance number, which lack interoperability, making cross-departmental data linkage inconsistent and inefficient. While workarounds such as fuzzy matching – which estimates connections based on name and address – can be used in some cases, they are insufficient for applications requiring precision, such as service delivery or personalised interventions.

The ONS (IDS) has adopted a spine-based matching approach, using three indices containing demographic, business and address information to link data sets. While this method improves accuracy, it remains a workaround rather than a long-term solution. For applications that require high confidence in identity verification, a standardised, government-wide personal identifier is necessary.

Other countries offer successful models for addressing this challenge. Finland and Estonia have adopted unique citizen identifiers, allowing seamless data linkage across public services. Estonia's X-Road system provides a secure, distributed platform for data-sharing between government systems, granting citizens control over who can access their data. This enhances trust and transparency while enabling efficient, digital-first government operations.

Recommendation: The UK should introduce a unique personal identifier to support accurate and efficient data linkage across public services. In the short term, ONS IDS's spine-based matching approach should be expanded, improving precision in linking demographic and administrative data sets. However, a universal personal identifier, integrated with the GOV.UK Wallet app and associated credentials, will be necessary to deliver personalised services at scale. This system should be designed with transparency and user control in mind, following models such as Estonia's X-Road, which enables secure, citizen-managed data-sharing while streamlining government services.

CREATING A COLLABORATIVE ECOSYSTEM: DATA BIOMES

To fully harness the potential of the NDL, a structured framework for cross-sectoral collaboration is needed. Many of the most consequential policy and technological challenges require coordinated data-sharing across departments, yet current efforts remain fragmented and reactive. A proactive, structured approach is required to foster data-driven innovation, enabling researchers, policymakers and businesses to develop real-world solutions to government challenges.

The NDL should facilitate the creation of Data Biomes – thematic, problem-focused data ecosystems that align with government priorities. These biomes will act as dedicated environments where structured data sets, technical expertise and real-world use cases come together to drive targeted innovation.

Recommendation: The NDL should establish Data Biomes to focus on high-impact thematic areas that align with government missions. Each biome should be chaired or co-chaired by relevant National Data Librarians, who will identify key challenges and opportunities, and publish specific data-driven problem statements open to participation from NDL-Reader-Pass holders in government, academia and industry. Innovators will gain access to relevant NDL data sets to develop and test solutions in a controlled sandbox environment.

Recommendation: At times when the government acts as the first customer for the solution, a streamlined procurement process should be put in place. This approach will enable developed technical solutions already built for a specific use case to be swiftly adopted in government, following a “[demos, not memos](#)” approach previously described by TBI. This efficient adoption of ready-made innovation would also accelerate public-sector use of AI.

In some cases, existing NDL data sets may be insufficient to develop a solution, requiring the generation of new, high-value data sets. Without a clear process for producing actionable data, innovation efforts may be slowed or misaligned with real-world needs. A Data Biome framework can systematically incentivise data production, ensuring that new data sets are created with specific applications in mind rather than as abstract assets with no immediate use. Data Biomes will also serve as testing grounds for cutting-edge techniques in AI, advanced analytics and digital governance. Rather than retrofitting existing data sets to work with new approaches, these biomes will embed the expertise, relationships and data infrastructure needed to rapidly experiment with and implement next-generation capabilities.

Recommendation: Data Biomes should act as signals for new data production, ensuring that government departments see direct value in improving data quality and availability. By focusing on highly targeted, actionable requests, the NDL will prevent unnecessary or premature data-set creation while ensuring that new data are structured for rapid deployment. This will eliminate the inefficiencies of collecting generic data

sets that require extensive transformation before they can be used and strengthen the motivation of government departments to prioritise high-quality data production.

05

Conclusion

The National Data Library is central to unlocking the UK's public-sector data and the applications and subsequent value these will enable. It will transform how data are used to drive innovation, improve public services and inform evidence-based policymaking across government, academia and industry. The framework for success is in place, but delivering results depends on decisive action and a clear vision now.

Immediate steps will show what the NDL can achieve, setting a foundation of credibility and momentum. Medium-term actions will scale its infrastructure, refine its governance and expand its capabilities. Long-term strategies will deliver a unified, transformative platform, ensuring the NDL becomes a cornerstone of the UK's data-driven future.

This is an opportunity the UK cannot afford to delay or fudge. The priorities are clear, the benefits are enormous and the tools are within reach. Now is the time for Britain to build.

06

Acknowledgements

The authors would like to thank the following experts for their input and feedback (while noting that contribution does not equal endorsement of every point made in the report).

Brendan Boyle, Fair Way Resolution

Rebecca Cosgriff, LifeArc

James Flemming, The Francis Crick Institute

Rosie French, Administrative Data Research UK

Emma Gordon, Administrative Data Research UK

Emily Jefferson, Health Data Research UK

Konstantinos Kaouras, Our Future Health

Edward Purchase, London Borough of Camden

Simon Ross, Stats NZ

Taj Sallamuddin, Information Governance Services

Cassie Smith, Health Data Research UK

Gavin Starks, Icebreaker One

Nick Swanson, Google DeepMind

Martin Waudby, London Borough of Camden

Stian Westlake, Economic and Social Research Council

We would also like to thank the organisers and attendees of roundtables and workshops that our authors have taken part in, including: the *National Data Library Design Lab* (organised by Connected by Data), *workshop on the UK National Data Library: Technical White Paper Challenge* (organised by the Wellcome Trust and the Economic and Social Research Council) and *AI Fringe 2025: How to Unlock Datasets for Growth (the National Data Library)* (organised by Startup Coalition).

Endnotes

- 1 <https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan>
- 2 <https://sciencesuperpower.substack.com/p/lets-get-real-about-britains-ai-status;>
<https://www.ukonward.com/reports/future-frontiers/>
- 3 <https://institute.global/insights/tech-and-digitalisation/act-fast-on-smart-data-to-unlock-gbp27-billion-in-economic-growth>
- 4 <https://www.adruk.org/news-publications/publications-reports/interim-evaluation-of-adr-uk-summary-report/>
- 5 <https://dareuk.org.uk/news-and-events/dare-uk-scientific-use-cases-workshop-report-published/>
- 6 <https://institute.global/insights/politics-and-governance/governing-in-the-age-of-ai-a-new-model-to-transform-the-state>
- 7 <https://assets.publishing.service.gov.uk/media/678f68b3f4ff8740d978864d/a-blueprint-for-modern-digital-government-print-ready.pdf>
- 8 UK National Data Library | Wellcome
- 9 <https://www.ukri.org/what-we-do/browse-our-areas-of-investment-and-support/administrative-data-research-uk-adr-uk/>
- 10 <https://www.adruk.org/news-publications/publications-reports/interim-evaluation-of-adr-uk-summary-report/>
- 11 <https://www.publictechnology.net/2020/10/28/business-and-industry/government-spend-20m-border-flow-contract-cia-backed-big-data-firm-palantir/>
- 12 <https://www.digitalhealth.net/2020/12/palantir-awarded-23m-deal-to-continue-work-on-nhs-covid-19-data-store/?utm%5Fsource=chatgpt.com>
- 13 <https://dareuk.org.uk/news-and-events/dare-uk-scientific-use-cases-workshop-report-published/>
- 14 <https://institute.global/insights/politics-and-governance/governing-in-the-age-of-ai-a-new-model-to-transform-the-state>
- 15 [https://www.climatechange.ai/dev/datagaps?showing=by-data set](https://www.climatechange.ai/dev/datagaps?showing=by-data%20set)
- 16 <https://russellgroup.ac.uk/news/research-intensive-universities-generate-nearly-38-billion-for-uk-economy>
- 17 UK National Data Library: Distributed Architecture for Research
- 18 <https://www.ons.gov.uk/economy/governmentpublicsectorandtaxes/researchanddevelopmentexpenditure/bulletins/businessenterpriseanddevelopment/>

2023

- 19 <https://www.gov.uk/government/publications/commercial-clinical-trials-in-the-uk-the-lord-oshaughnessy-review/commercial-clinical-trials-in-the-uk-the-lord-oshaughnessy-review-final-report>
- 20 <https://institute.global/insights/politics-and-governance/a-new-national-purpose-harnessing-data-for-health>
- 21 <https://www.gov.uk/government/news/teachers-to-get-more-trustworthy-ai-tech-as-generative-tools-learn-from-new-bank-of-lesson-plans-and-curriculums-helping-them-mark-homework-and-save>
- 22 <https://assets.ctfassets.net/75ila1cnaeh/68U0TqFbEjrucQVIEp3yl/3ef9a68c403ff650e27c87701ef593ea/The%5FEconomic%5FCase%5Ffor%5FAI-Enabled%5FEducation%5FFinal%5Fv2.pdf>
- 23 <https://www.ft.com/content/dd5515cc-e628-4e17-a4fd-1a10cc9f81e4>
- 24 See the Longitudinal Education Outcomes (LEO) data set, which links education outcomes, income, benefits and Inter-Departmental Business Register data <https://www.adruk.org/data-access/flagship-datasets/longitudinal-education-outcomes/>
- 25 <https://institute.global/insights/politics-and-governance/governing-in-the-age-of-ai-a-new-model-to-transform-the-state>
- 26 <https://www.gov.uk/guidance/national-underground-asset-register-nuar>
- 27 <https://reports.adruk.org/annual-report-2023-2024/our-data/new-and-emerging-datasets/ministry-of-justice-data-first-cross-justice-system-england-and-wales/>
- 28 <https://reports.adruk.org/annual-report-2023-2024/our-data/new-and-emerging-datasets/longitudinal-education-outcomes/>
- 29 <https://discover-now.co.uk/>
- 30 <https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/>
<https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/>
- 31 <https://www.ons.gov.uk/aboutus/transparencyandgovernance/datastrategy/relevantlegislation#:~:text=The%20Statistics%20and%20Registration%20Service,ONS%20is%20the%20Authority's%20exec>
- 32 <https://bills.parliament.uk/bills/3825>
- 33 <https://www.hdruk.ac.uk/helping-with-health-data/the-sudlow-review/>
- 34 <https://www.legislation.gov.uk/uksi/2002/1438/contents>
- 35 <https://nairrpilot.org/projects/awarded>
- 36 <https://e-estonia.com/solutions/x-road-interoperability-services/x-road/>
- 37 <https://www.onelondon.online/>
- 38 <https://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/disclosurecontrol/guidanceonintrudertesting>

- 39 <https://data.london.gov.uk/guidance/sharing-data/>
- 40 <https://ai.gov.uk/blogs/improving-legislative-drafting-with-lex>
- 41 <https://www.find-npd-data.education.gov.uk/>
- 42 <https://healthdatagateway.org/en>
- 43 <https://github.com/i-dot-ai/awesome-gov-datasets>
- 44 <https://odimply.org/files/case-study-denmark.pdf>
- 45 <https://takes.jamesomalley.co.uk/p/heres-the-plan-to-actually-liberate>
- 46 <https://theodi.org/insights/reports/how-an-ai-ready-national-data-library-would-help-uk-science/>
- 47 <https://arxiv.org/pdf/2205.03257>
- 48 <https://arxiv.org/abs/2211.01656>
- 49 <https://predictioncenter.org/>
- 50 <https://www.aria.org.uk/opportunity-spaces/scoping-our-planet/scoping-our-planet>
- 51 <https://unpredictablepatterns.substack.com/p/unpredictable-patterns-105-ais-with>
- 52 <https://www.ukri.org/news/ai-research-resource-funding-opportunity-launches/>
- 53 <https://www.researchdata.scot/accessing-data/researcher-access-service/ras-approval-pathway-pricing>
- 54 <https://www.aria.org.uk/wp-content/uploads/2024/03/ARIA-Standard-Grant-Agreement-V1.3.docx.pdf>
- 55 <https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/>

Follow us

facebook.com/instituteglobal

x.com/instituteGC

instagram.com/institutegc

General enquiries

info@institute.global

Copyright © February 2025 by the Tony Blair Institute for Global Change

All rights reserved. Citation, reproduction and or translation of this publication, in whole or in part, for educational or other non-commercial purposes is authorised provided the source is fully acknowledged Tony Blair Institute, trading as Tony Blair Institute for Global Change, is a company limited by guarantee registered in England and Wales (registered company number: 10505963) whose registered office is One Bartholomew Close, London, EC1A 7BL.