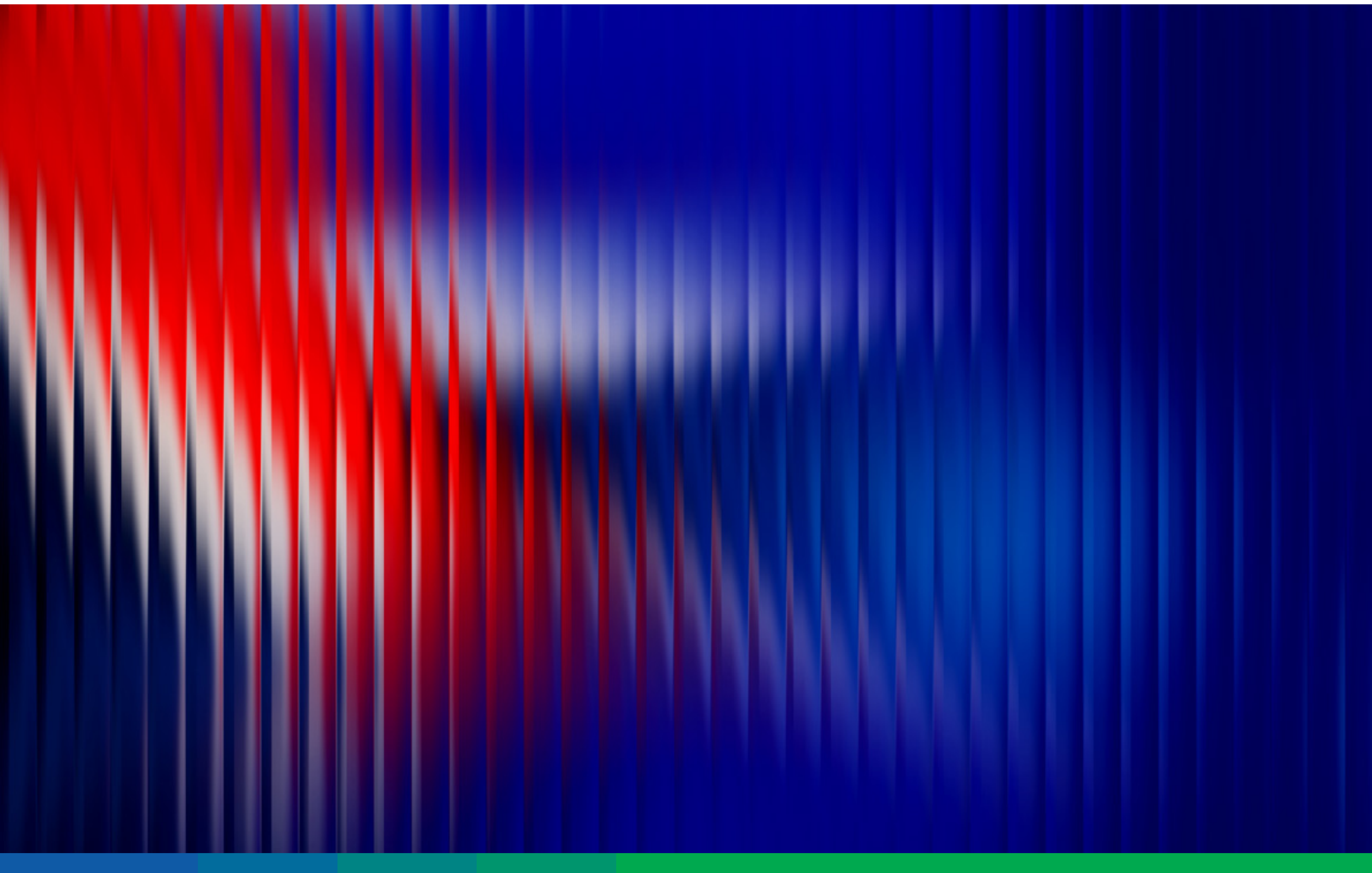
Research Report

Insights from UK councils on standards, readiness and reform to modernise public data for AI

Towards public sector modernisation with AI-ready data for search, machine learning and generative AI



Authors

This report was prepared jointly by the Open Data Institute and Nortal.

Neil Majithia
Researcher, ODI

Thomas Carey-Wilson
Researcher, ODI

Elena Simperl
Director of Research, ODI

Annabell Kuldmaa
Vice President, Government Transformation and Technology, Nortal

Priit Liivak
Chief Government Technology Officer, Nortal

Contributions

Kristi Kalmaru, Editorial lead
Melanie Rohde and **Annika Saar**, Design

How to cite this report

Simperl, E., Majithia, N., Carey-Wilson, T., Kuldmaa, A., & Liivak, P. (2025). Insights from UK councils on standards, readiness and reform to modernise public data for AI. Nortal & Open Data Institute.

Funding and independence

This research was commissioned and funded by Nortal. The Open Data Institute (ODI) co-developed the work, led research design and analysis, and retained full editorial independence.

Executive summary

The councils that modernise their approach to data will be the ones that modernise their services fastest.

Local authorities across the UK and Europe are rethinking their role: their datasets are no longer just about accountability and statutory reporting, but the raw material for automation, predictive modelling and chatbots that are reshaping public service delivery. AI-ready government data, including open data, is reducing operational costs, enabling data-driven policy and creating opportunities for private sector innovation, and in doing so, is becoming an economic catalyst. This positions local authorities not only as service providers but also as enablers of new markets, partnerships and citizen-centric digital ecosystems.

The Open Data Institute's AI-ready data framework has provided a strong baseline for assessing whether datasets are fit for algorithmic use. For councils, this matters because data quality and infrastructure determine whether they can forecast budgets accurately, prevent homelessness proactively or optimise transport networks efficiently.

At the same time, AI readiness is often used as an umbrella term for many related, but distinct concerns. AI is not monolithic. From the perspective of **continuous public sector modernisation**, different AI applications make very different demands on data. What works to power a search tool does not automatically work for a machine-learning pipeline; what works for predictive modelling may be entirely unsuitable for large language models (LLMs).

Councils need support to navigate these mismatches so they can make informed investments and ensure their infrastructure and procurement are fit for purpose.

To move towards an operational framework that acknowledges this diversity of AI techniques and use cases, ODI and Nortal ran a **structured programme of research** into the AI-readiness of local government data. We assessed ten high-impact use cases where councils are exploring or piloting AI. Six of them were based on interviews and closed datasets; four were analysed by our team using their open data sources. Each case was evaluated against dataset properties, metadata quality and supporting infrastructure, our three pillars of AI readiness. Together, these assessments provide a detailed picture of how close, or how far, local government data is from being usable in modern AI systems, including search, predictive analytics and LLM scenarios.

The findings show a mixed picture. Many councils still publish data mainly in formats designed for reporting rather than advanced technical use. Inconsistencies in formats, scarce metadata and limited infrastructure, such as missing APIs, search functionality or version control, remain barriers. Governance and skills gaps also contribute to the challenge. Yet there are encouraging signs: several councils are already taking proactive steps to improve data quality, strengthen infrastructure and experiment with new approaches. Building on these advances will be key to unlocking AI's full potential and ensuring public datasets contribute more fully to operational efficiency, service quality and policy effectiveness.

At the same time, our work shows a way forward. We propose an **expanded assessment framework:** building on ODI's three pillars (dataset properties, metadata and infrastructure), but interpreting them through three distinct readiness lenses:

Search readiness

Clean metadata, canonical identifiers and machine-readable schemas for discoverability.

Machine learning (ML) readiness

Structured, columnar data at scale, with identifiable imbalances and reproducible lineage.

Generative AI readiness

Richly annotated corpora, chunked unstructured text and APIs designed for retrieval and conversational use.

Taken together, these dimensions should not be seen as sequential stages but as a toolkit of options. Search readiness underpins transparency and efficiency; machine learning readiness supports foresight and planning; and generative readiness opens the door to new forms of citizen interaction. Each requires its own standards, governance and infrastructure, but together they give councils flexible pathways to modernise services and build trust.

This reframing moves the conversation from a simple ‚yes/no' on readiness to a more strategic choice about purpose. Crucially, our analysis also reveals a fundamental distinction between the AI-readiness requirements for secure, internal operational data and those for public, strategic data. Each readiness dimension aligns with a different major class of AI application, allowing councils to match ambitions with realistic steps toward capability.

This paper sets out our findings in detail. Each case study, from predicting rent arrears in Leeds to forecasting budgets in Ealing, from SAVVI's work on vulnerability data in Tameside to fire risk modelling in London, provides contextual background, a dataset overview, and an assessment of readiness. Common patterns emerge, but so too does a **roadmap for progress:** by modernising metadata, adopting standards and enabling secure data sharing, councils can move from "AI in theory" to AI that delivers on its promise – services that are efficient, transparent and trusted.

Table of contents

Introduction

AI readiness is often interpreted as a binary: data either is or is not "fit for AI." In practice, readiness is conditional on the intended use case. A dataset that works for one type of application may fail completely for another, with direct consequences for service delivery, budgets and public trust.

For example:

Search systems

One of the general obstacles to better data practice in UK local government is that information is still too often published in formats designed for reporting rather than reuse. Many councils still publish planning applications or deprivation statistics as PDFs. These satisfy reporting obligations but are not structured for efficient queries by address, date or category. As a result, officers spend hours manually extracting information for casework or policy analysis. By contrast, if the same data were tagged with clean metadata and canonical property identifiers, planning officers could answer resident queries in minutes, auditors could detect anomalies faster, and citizens could self-serve through online portals, reducing staff workload and improving transparency.

Predictive modelling

Councils often rely on spreadsheets of social care records to forecast demand. These lack the consistency required for reliable machine learning. If date formats vary across departments, or care packages are coded inconsistently, predictive models trained on this data can produce unreliable results. The outcome is costly pilots that fail to deliver and missed opportunities to anticipate rising demand. Conversely, structured and standardised records would allow finance teams to model future costs more accurately, supporting proactive interventions and more sustainable budgeting.

Conversational systems

Some authorities are beginning to test resident-facing chatbots for services such as bin collection, council tax or parking permits. These systems depend on underlying datasets that capture the necessary context, for example, service schedules that vary by postcode or local policy rules that affect eligibility. Without that structure, the chatbot risks giving incomplete or misleading answers, generating extra call centre demand rather than reducing it. By contrast, when datasets are enriched with metadata and governed by clear policy rules, conversational AI can provide accurate, auditable responses. This allows councils to cut routine call volumes, improve the resident experience and build trust in digital service channels.

These examples show why expanding beyond the simple one-size-fits-all notion of "AI readiness" is critical. **AI is not monolithic: search, predictive modelling and large language models (LLMs) each place different emphases on data, even if they also share important common ground. Good metadata, standardised formats, de-identification and strong governance are essential across all three, but the balance of what matters most shifts depending on the application.**

We therefore propose an expanded view. The foundation of this assessment combines the ODI's framework, built on three pillars of dataset, metadata and surrounding infrastructure, with an additional emphasis on governance and trust, and Nortal's perspective, which applies these pillars through the lens of different AI applications:

Search and conversational readiness

Clean metadata, canonical identifiers and machine-readable schemas to make data discoverable and usable in evolving interfaces.

Machine learning (ML) readiness

Structured, columnar data at scale, with identifiable imbalances and reproducible lineage.

Generative AI readiness

Richly annotated corpora, chunked unstructured text and APIs designed for retrieval.

This framework gives public bodies a more nuanced diagnostic tool, one that highlights crossovers as well as differences, helping them prioritise investments and plan transformation according to their needs.

The ODI–Nortal framework: readiness across search, machine learning and generative AI

Data is only as useful as the application it supports, which is why this framework looks at readiness through three distinct lenses: search, machine learning and generative AI.

There is a certain level of overlap between these dimensions, and the distribution is mostly intended to set priorities rather than setting focus. We aim to present councils and public bodies with a way of thinking – a baseline toolset adaptable to unique needs and environments, while recognising the distinct demands of different AI techniques rather than treating readiness as a single, linear maturity ladder.

Dimension	Search readiness	Machine learning readines	Generative AI readiness
Purpose	Enable AI-powered search to uncover patterns, anomalies, and semantic relationships within datasets, moving beyond keyword lookups to semantic exploration.	Provide stable, structured datasets for model training and retraining through reproducible pipelines.	Support generative systems in autonomously finding relevant datasets, interpreting context and reasoning across interactions between datasets.
Dataset	Clean, semantically consistent records; canonical identifiers; structured content and fields optimised for semantic indexing.	Structured/columnar formats; represen-tative samples; bias metadata.	Large and diverse corpora (structured + unstructured); contextual metadata linkages; explicit modelling of relationships between entities/events.
Metadata	Machine-readable de-scriptors that capture meaning and categories to power semantic que-ries and anomaly detec-tion.	Provenance, lineage, modality descriptors; explicit declarations of bias, imbalance and coverage to guide training.	Rich contextual metadata atta-ched at dataset and sub-data-set level to support grounding and reasoning. This includes shared vocabularies and onto-logies for the datasets.
Surrounding infrastructure	APIs and indexes that allow semantic queries, similarity search, anomaly detection and AI-driven pattern mining.	Reproducible pipe-lines; snapshotting and version control; automated quality checks; retraining triggers.	Embedding-ready APIs, vector databases, and orchestration layers for RAG; infrastructure that allows multi-dataset rea-soning.
Governance and trust	Standardised identifiers, transparency on update frequency, categorisation and labelling to ensure AI search results can be trusted.	Bias monitoring and explainability safe-guards; reproducible governance audits.	Policy-as-code for retention, privacy, and safety; safeguards against leakage; explicit inter-pretability rules for generative agents.

Readiness dimensions

When we talk about data readiness for AI, we are not describing a single ladder to climb. Instead, we are looking at three common technical architectures: search, machine learning and generative AI. From a technical point of view, these three are like apples and oranges – distinct approaches that prepare data in different ways to unlock different kinds of business value. And while we focus on these three in this framework, they are not the only dimensions. There are likely many more ways to think about readiness as AI capabilities evolve.

Search readiness

Search readiness is about enabling AI-powered search within datasets. Instead of scrolling through spreadsheets or PDFs, search-ready data allows AI to uncover patterns, anomalies and semantic relationships directly in the data. The value lies in faster analysis, better audits and more transparent access to information. Councils can use search-ready data to quickly flag inconsistencies in records, identify service demand spikes or group together related casework. It's about turning raw records into explorable asset.

Dorset Council's work in adult social care illustrates search readiness in practice. By aligning its datasets to standards such as ISO 8601 for dates and SNOMED-CT for health and care terminology, the council ensured records were consistent, machine-readable and interoperable across systems. This preparation enabled advanced search and exploration, allowing AI tools to flag anomalies, surface hidden patterns and make the data more usable for both operational casework and strategic planning.

Machine learning readiness

Machine learning readiness focuses on structuring data so it can train and retrain predictive models reliably. The emphasis is not just on having structured and representative datasets, but also on automated pipelines that refresh data, detect changes and update models as new information arrives. This ensures forecasts do not go stale and models remain aligned with reality.

Equally important is metadata that makes bias and imbalance visible. Declaring where data skews exist, documenting provenance and providing quality signals such as anomaly logs or summary statistics allows practitioners to trust model outputs and correct for distortions. Without such metadata, hidden weaknesses in the data surface later as model failures.

Bristol City Council applied machine learning to its Think Family Database, a multi-agency dataset that brings together education, social care, police and welfare records. Using this structured data, the council developed a predictive model to identify young people at risk of becoming NEET (Not in Education, Employment or Training). The project highlighted the importance of representative data and metadata: while the model could detect risk patterns, gaps in bias documentation and quality signals meant its outputs required careful interpretation before being acted upon.

Generative AI readiness

Generative AI readiness requires yet another kind of preparation. It involves providing context-rich, segmented datasets and infrastructure that supports generative AI and agent workloads. The aim is to let generative systems not only consume datasets but also interpret them, connect them with other data and in some cases identify relevant datasets on their own. The result is interaction and communication: citizen-facing assistants that can explain council tax rules, internal tools that help staff navigate complex policies or systems that provide grounded answers by reasoning across multiple datasets.

Different purposes, different outcomes

These readiness dimensions do not form a progression. One is not “higher” than the other. Each serves a different purpose, delivers a different type of value and requires its own investment in standards, metadata and infrastructure. The strategic task is to choose the readiness dimension that best fits the business challenge at hand, rather than treating them as sequential steps in a maturity model.

Findings on data, standards and readiness across UK councils

The following cases illustrate how councils across the UK are grappling with the realities of data standards and readiness. The findings reveal both progress and persistent gaps. Some councils are already demonstrating the value of AI-ready data, while others show how missing standards, weak metadata or limited infrastructure can hold back innovation. Taken together, these stories offer a candid snapshot of where local government stands today.

Aerial view of the London, England, Wirestock, Freepik

Dorset Council

Dorset Council:

Case-specific budget estimation in social care

Social care is one of the largest statutory responsibilities of local government, accounting for the majority of many councils' budgets. Forecasting individual and long-term needs is notoriously difficult: underestimates can leave vulnerable people without support, while overestimates lock up funds that could be used elsewhere. In 2010–11, Dorset Council commissioned Formulate, an AI tool trained and maintained on around 50 social care datasets to estimate the budget required for individual adults based on their needs profile. With over a decade in use, Formulate provides one of the clearest examples of AI a pplied to council finances, offering a test case for how data readiness supports responsible automation.

01

Observations

Dataset

Dorset has made progress in adopting standards such as ISO-8601 for dates and SNOMED-CT for health and social care, although legacy inconsistencies continue to affect interoperability. Bias is a recognised issue, reflecting Dorset's ageing population, and some personally identifiable data is retained because anonymisation would reduce accuracy – a deliberate trade-off between privacy and utility.

Metadata

Metadata is available in JSON, including supply chain information and technical details. However, assessments of bias and legal retention rules remain in documentation rather than attached to the datasets, reducing transparency and slowing audits.

Infrastructure

Dorset uses APIs and ETL pipelines for structured access, with timestamping for reproducibility. An internal interface supports discovery, but it is not yet a fully AI-centric data portal.

Readiness across search, ML and generative AI

Dorset's data infrastructure shows a mixed picture across different types of AI readiness. There is a solid foundation for AI-enabled search within datasets: data is structured and accessible within the council, and internal teams can surface information with reasonable reliability. On the machine learning side, Dorset has already demonstrated practical capacity. Structured datasets, reproducible access and explicit attention to bias have allowed predictive tools such as Formulate to deliver value in financial planning and resource allocation. This shows that Dorset's data environment is not static but capable of supporting automated pipelines and model retraining – a readiness that enables foresight rather than just insight.

Generative AI could add a new layer of accessibility to Dorset's financial data. Instead of requiring analysts or caseworkers to master technical tools or specialist terminology, a language model could let them query budget forecasts in plain English. Staff might ask questions such as "What are the projected costs for adults with complex needs over the next five years?" or "Where have budget estimates consistently overshot actual spend?" – and receive answers grounded directly in the underlying datasets. This kind of interaction would lower the barrier to insight, extend the use of financial planning data beyond specialists and create a more transparent decision-making environment.

Yet Dorset is not there yet. Contextual metadata is still sparse and not machine-readable, making it difficult for large language models to ground outputs or reason across datasets. The lack of segmentation, embeddings and orchestration layers means that generative systems cannot yet navigate Dorset's data landscape on their own. In this dimension, Dorset's readiness is at an early stage: promising datasets exist, but the infrastructure for generative use is not yet in place.

Recommendations

Dorset stands out nationally as a council actively prioritising AI-readiness. Its responsible AI framework, attention to ethics and bias and long-term experience with Formulate show the value of aligning governance with data infrastructure. The tool has already enabled Dorset to move from reactive budgeting toward proactive, evidence-based allocation – a major operational benefit given rising social care costs. To build on this progress, Dorset should embed missing metadata directly into datasets to streamline audits and reduce staff time and formalise reproducible pipelines with versioning and quality signals to guard against forecast errors. These improvements would not only strengthen current predictive tools but also open the door to LLM-powered assistants that could support caseworkers in budget planning and improve communication with families about resource allocation.

Corfe Castle, Dorset, Jim Champion from Southampton, UK. Creative Commons Attribution ShareAlike 2.0 www.ealing.gov.uk

Ealing Council:

Forecasting annual social care expenditure

Like all local authorities, the London Borough of Ealing faces growing pressures in adult social care, which represents a significant share of council budgets. To improve planning, Ealing developed the Adult Social Care Annual Expenditure Forecast, a tool designed to predict quarterly and annual spend by combining a financial model (estimating costs per individual) with an enrolment model (predicting how many people would join or leave social services). The system drew on data from the council's contracted social care management system. Although the model showed early promise, it was decommissioned within two years due to high costs and perceived inaccuracies . Council officers reported that forecasts frequently diverged from actual expenditure, undermining trust in the system. Maintaining the tool also required significant resourcing, from data preparation and pipeline upkeep to specialist oversight, which made it difficult to justify the investment given the limited reliability of the outputs. The decision to retire the model illustrates a critical lesson in machine learning readiness: without robust pipelines, transparent metadata and mechanisms to validate predictions over time, early pilots can falter even when the initial concept is sound.

02

Observations

Dataset

The tool was built on structured data from the social care management system, which staff considered consistent enough for internal use. However, the lack of adherence to common standards limited interoperability. In addition, anonymisation requirements, such as the exclusion of addresses, removed valuable context, which reduced the model's predictive accuracy. This highlights how councils face trade-offs between protecting privacy and preserving utility for decision-making.

Metadata

Metadata was largely absent. While teams described the data as "easy to use," the absence of machine-readable documentation meant that potential biases were not captured and reproducibility was harder to achieve. Without metadata, finance leaders lacked confidence in the model's outputs, which undermined trust in its value.

Infrastructure

Ealing maintained a capable technical environment, extracting case management data into local servers, avoiding the high costs of cloud services. However, the absence of pipelines for bias detection, metadata generation or version control meant the infrastructure was not optimised for advanced AI workflows, reducing the long-term sustainability of the model.

Readiness across search, ML and generative AI

Ealing's search readiness covers basic needs. Housing transaction data is detailed enough for officers to check payment histories and spot trends, but it is not prepared for AI-enabled search. Metadata is limited, bias is not clearly documented and there is little semantic structure. This means officers still depend on manual queries and dashboards rather than AI tools surfacing issues directly.

Ealing shows stronger progress on machine learning readiness. The social care expenditure forecasting model demonstrates that structured data can be used for predictive purposes, giving the council better foresight on budget demands. However, pipelines for ongoing retraining and bias monitoring are not fully in place. Current workflows rely on reproducible access to data but lack automated mechanisms to refresh models as new cases and costs emerge.

Recommendations

Ealing's forecasting project demonstrates a willingness to innovate and apply AI to one of local government's most complex financial challenges. Although the model was not sustained, it offered important insights into the conditions needed for predictive tools to deliver long-term value. The experience underlined that strong data foundations are essential for building trust in outputs and embedding AI in financial planning.

Ealing Town Hall (1888), P.g.champion Creative Commons Attribution 2.0 UK: England & Wales

Bristol City Counsil:

Identifying young people at risk of NEET

Supporting young people into education, employment or training is a critical statutory and social responsibility for local authorities. When young people become NEET (Not in Education, Employment, or Training), the consequences are significant: individuals face long-term disadvantages in income and wellbeing, while councils and wider public services incur higher costs through welfare support, social care and criminal justice. To intervene earlier, Bristol City Council developed a predictive NEET risk-scoring model, trained on the “Think Family Database” (TFD). This multi-agency dataset integrates records from education, social care, police and welfare agencies to flag vulnerable young people who share characteristics with historical NEET cohorts. Although direct access to the TFD is restricted for privacy reasons, Bristol has published an algorithmic transparency record describing the model, and aggregated ward-level data is available on its open data portal.

Observations

Dataset

The TFD integrates diverse data sources, from school attendance and attainment to child protection flags and household characteristics. Standard ward codes enable geographic linkage, but many metrics remain locally defined, making it difficult to compare across councils. Aggregation provides anonymity but reduces usefulness for individual-level risk modelling. The dataset's design makes imbalances visible, for example, higher NEET rates in disadvantaged wards, which is valuable for area-level analysis.

Metadata

Metadata is limited. While the transparency report and portal contain human-readable descriptions, they do not follow interoperable standards such as Croissant, nor are they attached directly to downloaded datasets. Crucial contextual details, like bias in outcomes (e.g., over-representation of males and certain ethnic groups in NEET statistics), are not embedded in metadata, making responsible reuse harder.

Infrastructure

Bristol's Open Data portal is modern and user-centric, offering search, filtering and dashboarding. An API is provided (OGC for geospatial data), though it is optimised for mapping rather than flexible tabular queries. However, version control is absent, meaning analysts cannot easily trace changes over time.

Readiness across search, ML and generative AI

Bristol's experience with the Think Family Education pilot shows how targeted datasets can be shaped into practical AI tools, even without perfect infrastructure. School attendance, attainment and vulnerability data were structured well enough to train decision-tree models that flagged pupils at risk of becoming NEET. This created clear operational value: safeguarding staff could identify vulnerabilities earlier and spend less time searching across disconnected records.

Yet, the datasets themselves still reveal limits in readiness for other potential high-value use cases. Metadata about socioeconomic bias and representativeness is only partially available, which reduces the transparency of predictions and complicates audits. Pipelines for refreshing and retraining models are not yet formalised, meaning that predictive accuracy depends on ad-hoc updates rather than systematic processes. While the pilot reduced information retrieval time and helped prioritise cases, the lack of consistent metadata and automated retraining leaves questions about sustainability over the long term.

Generative AI, meanwhile, is not within reach for this use case. The data is not segmented or enriched with contextual information that would allow a language model to explain predictions in plain language or connect risk factors across education and social care domains. Without this, conversational tools that could support teachers, social workers or families remain speculative rather than practical.

Recommendations

Bristol has shown that moderately prepared datasets can already deliver major improvements in frontline services when combined with predictive modelling. The NEET pilot demonstrates machine learning readiness translating into tangible outcomes: earlier interventions, reduced workload for staff and better prioritisation of support.

To deepen this progress, Bristol could focus on building automated pipelines that keep predictive models aligned with changing patterns in youth attendance and attainment. This step would reduce maintenance burdens and over time, the same foundations could support more advanced applications — including generative tools that explain risk scores or bridge communication gaps between schools, families and care providers.

Bristol Council House, Linda Bailey Creative Commons Attribution ShareAlike 2.0

Tameside Council (SAVVI):

Linking health and care data to identify vulnerability

Vulnerability often cuts across service boundaries: health, social care, housing and welfare data together shape how risks are identified and managed. Local authorities face major challenges in linking these datasets, both technically and legally, which makes it harder to spot risks early or plan joined-up interventions. The Scalable Approach to Vulnerability via Interoperability (SAVVI) project, developed by iStandUK in collaboration with Tameside Council, seeks to address this by creating a common standard for structuring and linking vulnerability-related data. By emphasising consistent metadata, provenance and governance, SAVVI aims to make datasets interoperable and trusted, giving councils a framework for vulnerability analysis that can be applied widely rather than through isolated pilots. Unlike tool-specific pilots, SAVVI is designed as a framework for interoperability that could be applied across councils.

04

Observations

Dataset

SAVVI sets out a common conceptual model for people, households and service relationships, making it easier to link datasets. This improves consistency but leaves responsibility for data quality and bias monitoring with individual councils. In the interview, it was emphasised that while interoperability is essential, it must go hand in hand with legality and governance to ensure trust in vulnerability models.

Metadata

Metadata is central to SAVVI's design, encouraging the capture of provenance and semantics through glossaries and shared definitions. The discussion highlighted that semantic clarity is often missing in council datasets, making this one of the framework's priorities. Adoption, however, remains uneven.

Infrastructure

Councils vary widely in digital maturity. SAVVI's long-term ambition is a secure portal with API access and query tools. However, many councils still prefer aggregated reports rather than working directly with linked datasets — a practical barrier to achieving full readiness.

Readiness across search, ML and generative AI

By promoting a consistent data model, SAVVI strengthens the ability of AI systems to search across previously disconnected records and highlight connections between services. This marks a clear improvement on siloed datasets where meaningful search or anomaly detection is more complicated.

The same consistency creates a foundation for predictive use. In principle, linked and standardised data could underpin models that flag households at risk of crisis earlier, helping authorities act before problems escalate. Yet because adoption of SAVVI is still patchy, and bias monitoring remains inconsistent, this potential is stronger in theory than in everyday practice.

Generative applications remain the furthest away. Councils still lack the contextual metadata, APIs and retrieval infrastructure that would allow large language models to explain risks conversationally or connect household data across services in real time. SAVVI points in this direction, but the groundwork for such tools has yet to be laid.

Recommendations

SAVVI demonstrates that AI readiness is not only about individual datasets but about how they connect across services. By promoting interoperability, the framework gives councils the potential to identify vulnerable households earlier, reduce duplication and plan interventions with greater confidence. These outcomes translate directly into financial efficiency, reduced pressure on frontline teams and better long-term wellbeing for residents. The emphasis on metadata and lineage is particularly valuable for building trust in predictive models and, over time, generative tools that could help staff explain risks and decisions more clearly to families. The key challenge now is adoption: embedding SAVVI standards into governance and procurement, investing in consistent metadata practices and piloting API-enabled access. If councils follow through, they can move from fragmented silos to a more mature data ecosystem that supports preventative, cost-effective services.

Tameside Council Offices, Steven Haslington,
Creative Commons Attribution-Share Alike 2.0

Essex County Council

Essex County Council:

Tracking power use to cut costs and emissions

Energy consumption is closely tied to both sustainability and economic development. For councils, understanding how power flows through public buildings and services is not just about cutting costs, it reveals hidden inefficiencies, tracks progress toward climate and net-zero goals, and signals how well communities cope with demand and disruption. Essex County Council (ECC) works with a third-party provider to access the Bureau Monthly Cost and Consumption Matrix, a dataset covering power usage and costs across public buildings in the county. This dataset could underpin analysis of pollution levels, sustainability planning and infrastructure resilience, but its value depends on whether it is AI-ready.

05

Observations

Dataset

The dataset reports monthly consumption (kWh) and associated costs (£) for each building ("site"). However, ECC and the service provider use different site codes, making linkage with other datasets difficult. Inconsistencies and duplication were reported, such as the same building appearing under multiple identifiers. Without alignment to standards like UPRNs/USRNs, the dataset is hard to integrate into wider analysis.

Metadata

Metadata is almost non-existent. Aside from the dataset name and a basic description on the portal, users receive no machine-readable metadata, summary statistics or contextual information. Errors and anomalies are not flagged, leaving analysts to identify them manually.

Infrastructure

The dataset is provided through an invite-only portal that allows one-click downloads, which analysts find easy to use. However, there is no API access, no integrated quality checks, and no versioning beyond monthly file updates. Users are not automatically alerted when anomalies are corrected, which can delay or undermine analysis.

Readiness across search, ML and generative AI

Essex's energy data is structured but poorly prepared for AI-enabled search. Although the dataset can be downloaded easily, the lack of metadata and inconsistent identifiers means it is difficult to discover, interpret or link with other council datasets. As a result, insight remains locked behind manual reports rather than being surfaced interactively by AI systems.

The dataset could in principle support predictive modelling of energy demand and cost trends, but machine learning readiness is held back by the lack of automated pipelines and transparent metadata. Forecasting tools would need reproducible access to updated records, and without bias checks, models could misrepresent real consumption patterns.

Generative AI applications are currently out of reach. An example of such use case has been explored by Justice, Vakaj, and Dridi (2024), who developed EnergyChat, a LangChain-powered chatbot that demonstrates how generative AI, when combined with knowledge retrieval, can bridge these gaps by explaining energy usage trends in plain language and offering personalized, context-aware recommendations for sustainable consumption. Essex's energy data is missing contextual metadata and APIs, and the proprietary control of the dataset prevents embedding-based retrieval or conversational access. Generative systems cannot yet explain usage trends in plain language or link energy consumption to wider environmental datasets.

Recommendations

Essex's case highlights both the promise and the pitfalls of third-party data services. On the one hand, structured consumption records create a potential foundation for sustainability planning, pollution monitoring, and cost reduction. On the other, limited metadata, proprietary formats, and lack of open infrastructure constrain AI readiness. To unlock value, ECC would need to secure access in open, machine-readable formats; embed metadata on provenance, bias, and representativeness; and develop pipelines for reproducible updates. These improvements would enable more accurate forecasting, more transparent audits, and eventually generative tools to help staff and citizens understand energy use in the context of climate goals.

Headquarters of Essex County Council, Richard Kelly

Lancashire County Council:

Forecasting traffic to plan greener transport

Traffic patterns shape the daily lives of residents and the long-term sustainability of infrastructure. For local authorities, accurate forecasting can reduce congestion, support cycling and walking initiatives, and underpin investment in greener transport networks. Lancashire County Council (LCC) uses AI-powered cameras to count vehicles, cyclists and pedestrians across the county, feeding into its sustainable travel planning. This data could provide a powerful basis for forecasting and mitigation. However, as with many local authority datasets, their value depends on how far they meet the conditions of AI-readiness.

06

Observations

Dataset

The dataset benefits from ISO 8601 timestamps and unique identifiers, making linkage technically possible with other Vivacity deployments. But LCC's internal naming conventions for "countlines" (the points where traffic is recorded) are inconsistent and increasingly opaque as more cameras are added. Analysts often need to cross-check with separate mapping files, limiting efficiency and raising the risk of errors.

Metadata

Machine-readable metadata is absent. While the dataset includes a valuable "availability" column indicating the reliability of each measurement, there are no summary statistics, contextual details or information on provenance. Analysts must rely on internal knowledge and contracts, leaving gaps in transparency and interoperability.

Infrastructure

Data is accessed through Vivacity's online portal, which provides basic dashboards and download functionality. Although an API exists, council policy prohibits its use, restricting analysts to static files. This prevents automation and limits the dataset's scalability for AI applications.

Readiness across search, ML and generative AI

Lancashire's traffic data is well suited for interactive search. Officers can already query flows across locations and times, and the structured counts mean AI systems could help compare patterns between modes of travel or surface hotspots of congestion. What is missing is semantic metadata to allow richer natural-language search, such as asking which areas saw the biggest shift from car to cycling over a season or connecting usage patterns with weather or event data.

The dataset is also a strong candidate for predictive modelling. Continuous, structured records are exactly the kind of input needed to train forecasting models. But machine learning readiness depends on automated pipelines and bias monitoring, both of which are limited. Without systematic retraining, models risk becoming less accurate as travel behaviours change. And without documented metadata on geographic and demographic bias, forecasts could systematically under-represent certain communities or modes of travel.

Recommendations

Lancashire's use of AI-powered traffic monitoring shows clear progress toward data-driven planning for greener transport. The datasets provide a strong baseline for both interactive analysis and predictive modelling, and the council has already demonstrated the potential of AI-enabled infrastructure in practice. To strengthen readiness, Lancashire should focus on adding semantic metadata to improve search and ensure coherence in conventions. These improvements would make forecasts more accurate, support transparent evaluation, and lay the groundwork for future generative tools that could help staff and citizens explore traffic scenarios in plain language.

County Hall, Preston, UK Francis Franklin, Creative Commons Attribution-Share Alike 4.0 International

GREATER
LONDON
AUTHORITY

Greater London Authority:

Using fire data to predict and prevent outbreaks

Fire incidents are not only matters of public safety but also indicators of wider risks tied to housing quality, urban planning and community resilience. For regional authorities, predicting and preventing fires can save lives, reduce costs for emergency services and inform broader policy on housing, health and infrastructure.

The London Fire Brigade (LFB), supported by the Greater London Authority (GLA), manages one of the most comprehensive fire incident datasets in the UK. Its structured records, available on the London Datastore, cover every incident attended since 2009. With over 70 coded variables and tens of thousands of entries, this dataset provides a rich foundation for predictive modelling. The GLA's case highlights how strong open data practices can bring authorities closer to AI-readiness, while also revealing gaps between structured, public datasets and the unstructured, internal data that remains largely untapped.

Observations

Dataset

The LFB's structured dataset contains over 70 coded columns spanning date, time, borough, ward, property type, incident type, number of fire engines and personnel deployed. It adheres to strong standards, including ISO-compatible time formats and standard geographical codes. Class imbalance, such as the prevalence of false alarms compared to serious fires, is visible and therefore manageable. Unlike many local authority datasets, no de-identification is needed: the dataset is designed for public release. By contrast, the internal dataset of 37,000 free-text investigator reports since 2000 offers valuable qualitative detail but is unstructured, inconsistent and initially resembled a "data swamp."

Metadata

Metadata is a notable strength. Each dataset on the London Datastore is accompanied by a dedicated data dictionary spreadsheet that documents every field and its codes, ensuring analysts can interpret data consistently. However, this metadata is not machine-readable, meaning AI tools cannot automatically ingest the schema. Improving this would enable smoother integration into AI pipelines.

Infrastructure

The London Datastore exemplifies best practice in public sector data portals. Users can explore the fire dataset through embedded Power BI dashboards before downloading the full files in CSV format. Licensing under the UK Open Government Licence (OGL v2) supports transparent reuse. Two weaknesses limit AI-readiness: the absence of an API for programmatic access and the lack of formal version control to track changes across monthly updates. Both are highlighted in ODI's framework as best practice for AI-ready infrastructure.

Readiness across search, ML and generative AI

The fire dataset is well prepared for interactive and AI-enabled search. Structured coding allows users to query incidents by category, time or location, and AI systems could support pattern recognition such as hotspots of electrical fires or seasonal risks. Richer semantic metadata always benefits such usage and would allow users to run natural language queries such as: "show me patterns of fires linked to overcrowded housing."

For predictive modelling, the dataset is one of the strongest examples among UK authorities. Longitudinal consistency, structured variables and reproducible access via the Datastore make it a prime candidate for training and retraining models. Such models could forecast likely fire risks by borough, housing type or season, enabling more targeted prevention campaigns and resource allocation. The key gaps lie in bias documentation and integration: without metadata on reporting imbalances, models may over- or understate risks in particular communities, and without linkage to housing and social datasets, predictions cannot reflect the full social context of fire risks.

Generative AI readiness is less advanced. While the dataset is openly accessible, it lacks the contextual annotations and segmentation required for large language models to explain risks or connect fire patterns to wider policy questions.

Recommendations

The LFB case underscores the dual challenge for public bodies: maximising the potential of structured open datasets while unlocking value from messy, unstructured archives. The public incident dataset is highly AI-ready and could underpin predictive modelling for incident hotspots, resource planning and borough-level risk profiles. In contrast, the 37,000 unstructured investigator reports remain underused, despite containing insights into emerging risks such as grease build-up in restaurant extraction systems.

For GLA and LFB, the next step is twofold: (1) continue publishing high-quality, structured open data with improvements to metadata and API access, and (2) develop internal capability to transform legacy unstructured archives into AI-ready corpora for secure use. Taken together, this strategy would not only enhance fire prevention but also set a benchmark for how public authorities can manage both open and internal data assets for maximum societal value.

The interior helical staircase of City Hall, Colin
Creative Commons Attribution-Share Alike 4.0 International

Kent County Council:

Using predictive analytics to prevent homelessness

Homelessness prevention is one of the most high-stakes applications of predictive analytics in local government. The ability to identify vulnerable households three to six months before a crisis occurs can reduce hardship, cut the cost of emergency interventions and strengthen community resilience. Kent County Council (KCC) and Maidstone Borough Council (MBC) are piloting the "One View" platform to integrate financial, socio-economic and Equality Act characteristics into a predictive model for early intervention.

While the operational dataset that powers One View is not public, it draws on highly sensitive records such as council tax arrears, housing benefits and social care case notes, KCC does publish related strategic data. This includes a "Financial Hardship Toolkit" and reports on deprivation, benefits and socio-economic indicators. These public assets provide a useful proxy for evaluating AI-readiness and illustrate the sharp divide between the requirements of operational AI systems and the state of publicly available data.

08

Observations

Dataset

The public data is fragmented across reports and spreadsheets, with no single, coherent dataset. While some components, such as the Index of Multiple Deprivation (IMD), adhere to national standards, others lack semantic or logical consistency. Formats are a major weakness: critical indicators are published in PDFs and XLSX files, requiring manual extraction before analysis. Strengths are limited to the fact that the data is anonymised and designed to highlight imbalances across districts, which is valuable for strategic planning but not operational use.

Metadata

Metadata is not available in a structured form. Definitions, sources (e.g., ONS) and methodological notes are buried in the narrative of PDF reports. No machine-readable metadata, no technical specifications and no attached legal licensing information accompany the downloads. Although the council's website references the Open Government Licence, this is not directly applied at file level, representing a failure to meet ODI standards.

Infrastructure

The data is hosted on a generic "Facts and figures about Kent" webpage rather than a dedicated data portal. The page offers only hyperlinks to files, with no search, filtering, visualisation or analytical tools. A separate "Open Kent" portal exists but does not host these datasets. There seems to be no API access or version control, which means developers using this data cannot track updates or changes across reporting cycles.

Readiness across search, ML and generative AI

Kent's public data demonstrates weak search readiness. Without metadata, consistent identifiers or a modern portal, discovery is manual and inefficient. AI-enabled search across datasets is impossible, as systems cannot interpret structure or semantics beyond raw tables and reports.

Machine learning readiness is also very low. While indicators such as deprivation scores and unemployment rates could, in principle, underpin predictive models, their current formats (PDFs, scattered spreadsheets) make programmatic analysis laborious. The lack of pipelines and reproducibility further limits their use for training and retraining.

For generative AI, readiness is effectively non-existent. Without structured metadata, segmentation or APIs, these datasets cannot support conversational tools to answer practical questions such as "Which districts are showing early signs of rising homelessness risk?" or "How does benefit uptake correlate with arrears?"

Recommendations

The operational dataset powering One View is highly sensitive, real-time and multi-agency. Its purpose is to enable preventative action by identifying specific households at risk. The public data published by KCC, by contrast, is aggregated, thematic and designed for strategic reporting. These two data classes, operational vs. strategic, have fundamentally different characteristics and AI-readiness requirements.

For councils, this distinction highlights the need for two parallel data strategies:

1. Public strategic data should be made AI-ready by converting PDFs to open, machine-readable formats (e.g., CSV or Parquet), embedding metadata (e.g., Croissant or JSON-LD) and publishing via APIs. This would support transparency, research and long-term policy evaluation.

2. Operational data requires robust governance, pseudonymisation and controlled access to enable high-impact predictive tools like One View, while safeguarding citizens' privacy.

Kent's experience underscores that predictive analytics for prevention depends not only on advanced platforms but also on a clear, dual-track approach to data management, one for openness and accountability, the other for secure operational use.

Arcade at Ramsgate Harbour, Keith Edkins, Creative Commons Attribution-Share Alike 2.0 Generic

Leeds City Council:

Predicting rent arrears to stop debt spirals

Rent arrears remain one of the most pressing challenges for local housing teams. Traditional arrears management is reactive, with officers contacting tenants only after balances have already built up. RentSense, an AI-powered tool adopted by councils including Camden, Lambeth and Epping Forest, reverses this approach. By analysing rent-transaction histories, it predicts which tenants are most at risk of falling behind, allowing officers to intervene earlier and prevent spirals of debt.

The core datasets that drive RentSense, tenant-level rent accounts and payment histories, are highly sensitive and not publicly available. For this evaluation, we use Leeds City Council's "property lets" dataset from the Data Mill North portal as a high-quality proxy. Although it describes housing allocations rather than arrears, its granularity, structured format and features such as applicant priority and waiting times make it an excellent analogue for the kinds of variables an arrears prediction model would require.

09

Observations

Dataset

The dataset is published in CSV, a preferred, universally machine-readable format, making it more AI-ready than PDFs or XLSX files. Its schema is stable across years and uses consistent codes for applicant priority (A, B, C), enabling comparability over time. Each record includes property type, ward, applicant priority, number of bids received and rehousing time — features that mirror the kinds of risk factors operational models use. While class imbalance exists (e.g., some wards or property types are over-represented), it is identifiable and therefore manageable. The main weakness is that the dataset uses council-specific codes rather than national property standards.

Metadata

Metadata provision is relatively strong. A human-readable data dictionary explains key fields and business terms like "AP Priority." However, metadata is not machine-readable, and important elements such as summary statistics or bias statements are missing. This limits discoverability and slows exploratory data analysis.

Infrastructure

The dataset is hosted on Data Mill North, a user-centric portal that goes beyond static council webpages by providing search and download functionality. However, advanced AI-ready features are absent: there is no API for programmatic access, no facility for incremental queries, and no version control to track changes across updates. Analysts can only download the latest file, creating risks for reproducibility.

Readiness across search, ML and generative AI

Leeds demonstrates reasonable search readiness: the dataset is documented, discoverable, and provided in a usable format. But the absence of machine-readable metadata and standardised codes reduces interoperability and restricts AI systems from offering richer, query-driven insights.

The data shows strong machine learning readiness. Structured, consistent records with coded features can be readily used for modelling, and imbalances are visible enough to be managed. Adjustments for local codes and schema alignment would be needed, but the dataset is close to analysis-ready.

Generative AI readiness remains weak. Without APIs, structured metadata, or segmentation, the dataset cannot support natural language queries such as "show me priority-C applicants waiting more than 12 months in Armley ward." Nor could it underpin conversational tools for officers or citizens.

Recommendations

As with homelessness prevention in Kent, this case illustrates the divide between operational data (sensitive, transactional records powering tools like RentSense) and strategic data (aggregated or thematic datasets made public for transparency). The Leeds property lets dataset is high-quality for a public resource and provides an excellent proxy for arrears-related modelling. However, improving its AI-readiness, by converting council codes to national standards, embedding machine-readable metadata and adding API and version control support, would mainly benefit research and transparency rather than operational arrears prevention.

For councils, the key insight is that expectations for AI-readiness must differ across data classes. Operational datasets need secure, real-time integration with robust governance, while strategic open datasets must prioritise accessibility, interoperability and reproducibility. Recognising this duality is essential if housing services are to balance innovation in AI-driven arrears prevention with public accountability and trust.

Leeds Civic Hall owl (and clock), Chemical Engineer,
Creative Commons Attribution-Share Alike 4.0 International

Ealing Council:

Forecasting budgets with AI

Local authorities face immense difficulty in accurately forecasting annual expenditure due to the heterogeneous nature of service enrolment, fluctuating costs and variable rates at which individuals enter and leave care systems. Expenditure on areas such as adult social care, children's services and homelessness can swing unpredictably from year to year, making accurate forecasting critical for responsible financial planning. As Ealing Council's pilot has shown, AI models can improve April-to-April budget predictions, helping leaders allocate resources more effectively and reduce financial risk.

At the foundation of any such initiative is the "Local authority revenue expenditure and financing England" statistical series, published annually by the Department for Levelling Up, Housing and Communities (DLUHC). This dataset provides a standardised, authoritative picture of every English council's expenditure, covering outturn (actuals) and budgeted forecasts across all service areas. While designed for reporting and accountability, its structure and scope make it a valuable training and benchmarking resource for AI-driven financial forecasting.

Observations

Dataset

The dataset is highly reliable, reflecting mandatory submissions from all local authorities. It adheres to UK government accounting standards, ensuring semantic and logical consistency across entries. Expenditure imbalances (e.g., the dominance of social care) are clearly visible, making them usable features rather than hidden flaws. Privacy is not a concern since data is aggregated at the authority level. Weaknesses lie mainly in the file formats: XLSX and CSV are functional but not optimised for large-scale machine learning, where columnar formats like Apache Parquet would perform better.

Metadata

Documentation exists in the form of "Technical Notes," but it is detached from the dataset itself and not machine-readable. While definitions of expenditure categories are clear and high-level summary statistics are provided, critical gaps remain: there is no structured metadata file, no declarations of bias and no information on the handling of synthetic data. Licensing is provided under the Open Government Licence, but again not in an attached, machine-readable format.

Infrastructure

The dataset is published via data.gov.uk, the UK's central repository. This guarantees discoverability but limits usability: the portal is a file store, not a practitioner-oriented platform. Preview tools exist, but there are no integrated analytical functions or visualisations. API access is possible but poorly optimised for granular queries (e.g., service-level spending in a single council). Versioning is functional, datasets are released by year and release number, but does not provide machine-readable change logs or diffs, which are essential for reproducible AI pipelines.

Readiness across search, ML and generative AI

Ealing's pilot shows how a national dataset, designed for reporting and accountability, can be repurposed for predictive use. The consistency of the DLUHC expenditure series — with every authority submitting to the same schema each year — creates a rare opportunity for training models on a reliable, comparable baseline. This is a strong foundation for machine learning, even if the formats and infrastructure create friction in practice.

At the same time, gaps in metadata and delivery limit both transparency and automation. Definitions are clear in the technical notes, but without structured, machine-readable descriptors, AI systems cannot easily parse categories, trace provenance, or surface anomalies. For frontline users, this means the data is searchable by humans but not optimised for AI-enabled queries.

Looking further ahead, conversational tools that could explain spending patterns or compare trends across councils remain out of reach. Without APIs designed for granular queries, and without embedded metadata that would let large language models ground their answers, the dataset cannot yet support generative AI applications.

Recommendations

The financial dataset sits at the heart of local government planning, and its quality is unquestionable for reporting and human analysis. But when judged against our AI-readiness framework, it falls short of enabling automated, AI-driven use. The challenge is not the data itself, which is standardised, comprehensive and authoritative, but the surrounding infrastructure.

To make this dataset truly AI-ready, central government should prioritise:

- publishing in modern, columnar formats (e.g., Parquet)
- embedding machine-readable metadata (e.g., JSON-LD via Croissant) directly into files
- improving API endpoints for granular, programmatic access
- implementing machine-readable version control logs

The lesson for councils is clear: AI-readiness often requires modernising the delivery of data, not the information itself. By addressing these infrastructure gaps, national datasets like this could power not only compliance reporting but also predictive tools that help councils forecast costs, allocate resources and build financial resilience.

Ealing Council building, Philafrenzy, Creative Commons Attribution-Share Alike 4.0 International

Discussion: AI-ready data requires AI-ready decision making

Across these ten case studies, there are more commonalities than differences between local authorities and the AI-readiness of their data. For example, datasets are almost all accessible via data portals, while metadata is rare (with machine-readable metadata non-existent). The most visible differences were in the extent of datasets' adherence to standards, where Essex County Council struggled in data linkage because of a mismatch in site-code standards and Dorset's adoption of SNOMED-CT allows its datasets to be applicable across health and social care contexts.

Notably, the findings of each AI-readiness assessment criterion represent a decision, a choice that was made at one point in the lifecycle of the dataset, whether in its design, procurement, completion, iteration, analysis or usage in AI contexts. A decision will have been made to create a data portal or to adopt a standard or to not document contextual information as metadata. Some of these choices enhanced the AI-readiness of datasets, others harmed it, and while each may have been made by councils on the basis of short-term OKRs, monetary costs, technical knowledge (or a lack thereof) or any other justification, it is important to note that these decisions are what propagate into making a council's data AI-ready or not and, by extension, directly affects the services they can provide to their residents, now and in the future.

Such decisions are not easy to make. We hope that this research, including both the framework it introduces and the vignettes it showcases, can provide a structured evidence base for leadership, especially in resource-constrained, under-pressure organisations like UK local authorities, to make sure they navigate decisions correctly and build data (in addition to products and services built on it) that is at the cutting edge, ready for AI.

Conclusions: A toolkit for local authorities

Search readiness, machine learning readiness and generative AI readiness represent distinct dimensions of maturity, but they are not entirely parallel. All three depend on a strong data culture, shared governance practices and consistent standards.

Search readiness, a theme that has been on the government agenda since the early days of open data and the creation of the ODI more than a decade ago, remains foundational. Machine learning readiness builds on this foundation, enabling predictive models that support foresight and planning. Generative AI readiness is newer and less mature, but it highlights emerging opportunities to transform how councils interact with citizens.

We observe local authorities progressing unevenly across these dimensions. Many have made strides in search readiness, though challenges with metadata and discoverability persist. A smaller number are showing progress in predictive modelling pilots, while generative AI readiness is only beginning to be explored. Some councils may focus mainly on predictive ML, and that is entirely valid, but they should also be aware of the added value that generative approaches will bring.

This uneven progress is not a weakness but a reflection of choice and context. Councils can adopt the readiness dimensions that best fit their priorities while recognising that over time, all three will reinforce one another. Search readiness underpins transparency and accountability, predictive ML enhances efficiency and planning, generative AI promises new ways to engage communities and improve trust.

The UK has already laid important groundwork, from open government data initiatives to legislation promoting data sharing and analytics, and councils are building on this legacy. The future of local government is not only about catching up to technological change but about leading with it: creating services that are efficient, transparent and trusted, and communities that are stronger, more resilient and more connected.

Bibliography

Open Data Institute (ODI). (2012). Open data, open society: A research project. London: ODI.

Open Data Institute (ODI). (2025). A framework for AI-ready data. Available at: https://theodi.org/insights/reports/a-framework-for-ai-ready-data/

Majithia, N., Carey-Wilson, T., & Simperl, E. (2025). Local authorities, data, and AI: Phase 1 short paper. Open Data Institute & Nortal.

Nortal. (n.d.). AI Use Case Repository: Social benefits, support and protection. Nortal. Available at: https://nortal.com/ai-use-cases-public-sector

Local Government Association (LGA). (2025). State of digital local government. Available at: https://www.local.gov.uk/publications/state-digital-local-government

Local Government Association (LGA). (2024). Local government facts and figures. Available at: https://lgiu.org/resources/local-government-facts-and-figures/local-government-facts-and-figures-england/#section-3

Department for Levelling Up, Housing and Communities (DLUHC). (2024). Local authority revenue expenditure and financing England: statistical release. Available at: https://www.gov.uk/government/collections/local-authority-revenue-expenditure-and-financing

Greater London Authority (GLA). (2024). London Datastore. Available at: https://data.london.gov.uk/

UK Government. (2015). Open Government Licence v3.0 (OGL). Available at: https://www.nationalarchives.gov.uk/do%pen-government-licence/version/3/

Justice, K., Vakaj, E., & Dridi, A. (2024, July). AI insights: Unveiling UK energy consumption with LangChain-powered chatbots. Available at: https://ec-3.org/publications/conferences/EC32024/papers/EC32024_213.pdf

Appendix: Methods and supporting notes

Research design and methodology

This study was conducted jointly by the Open Data Institute (ODI) and Nortal between March and August 2025. The research design combined qualitative and documentary methods:

- Case study selection: Ten local authority use cases were chosen to represent a diversity of service areas (social care, housing, transport, fire safety, energy and vulnerability data) and to illustrate different stages of AI-readiness. Selection was based on desk research and consultation with sector experts.
- Interviews: Semi-structured interviews were carried out with council staff, technical leads and programme partners. Participants gave informed consent and were assured anonymity unless they requested attribution.
- Document and data analysis: Publicly available datasets, portals and technical documentation were reviewed against the ODI framework's three pillars (dataset properties, metadata, infrastructure) and Nortal's extended readiness dimensions (search, machine learning and generative AI readiness).
- Analytical rubric: Each case was assessed using a structured scoring template covering dataset properties, metadata provision, surrounding infrastructure and governance practices. Findings were synthesised into narrative case studies with readiness conclusions.

Use of data standards

In assessing AI-readiness, we noted the application (or absence) of established standards:

- **SNOMED CT®** – clinical terminology for health and social care, referenced in case studies such as Dorset Council. SNOMED CT® is a registered trademark of SNOMED International, and use of SNOMED content is subject to its licensing terms.
- **UPRN/USRN identifiers** – Unique Property Reference Numbers and Unique Street Reference Numbers, considered best practice for linking housing, planning, and infrastructure datasets.
- **ISO 8601 date/time formats** – highlighted as important for interoperability across datasets and predictive pipelines.

Where standards were applied, they were documented to support reproducibility. Where absent, their omission was noted as a limitation for discoverability, bias detection or interoperability.

Ethical considerations

- All interviews and case contributions were conducted with informed consent.
- No personal or sensitive data are published. Operational datasets remained under the governance of their respective controllers.
- Examples are anonymised or aggregated to avoid disclosure of individual-level information.
- The project followed the principles of the UK GDPR and the Data Protection Act 2018.

Limitations of the research

- **Sample size:** Ten case studies cannot represent the full spectrum of UK local authorities.
- **Heterogeneity:** Councils use different platforms and infrastructures, which limits comparability.
- **Data quality:** Inconsistent standards, lack of machine-readable metadata and limited version control affected the reproducibility of some assessments.
- **Rapid change:** AI tools and standards are evolving quickly. Findings reflect the state of practice in mid-2025.

Acknowledgements

We thank participating councils and interviewees for their time and insights. We also acknowledge colleagues at ODI and Nortal who provided methodological input and peer review. Any errors remain the responsibility of the authors.

nortal.com