

Smart Data

Response to the Department for Business, Innovation, Science and Trade's multi-sector call for evidence

Smart Data in rail

What opportunities and benefits would each use case¹ bring to the rail sector? For whom, and why?

The second rail use case presents a clearer public value case. The first one also has potential, particularly during disruption, but would require timely access to and use of operational data and better coordination rather than simply more customer data.

Rail use case 1: Supporting improved journey planning information across the rail system

Consistent information across the rail system on crowding and seat availability, stations and platforms, facilities, accessibility, service conditions and connections, updated promptly as conditions change, could help passengers choose between alternatives and reduce uncertainty. It could also give frontline staff a more coherent picture during disruption. Existing open services are useful in understanding incidents and sometimes provide more up to date information than is available to staff. A joined-up service could therefore improve confidence if it can process and present changing information quickly enough.

The service should also draw on behavioural evidence about how people travel and respond to disruption. Data on journeys taken, passenger demographics and accessibility needs, the purpose or time-sensitivity of a journey can help determine what information passengers need, when they need it and how it should be presented.

Transport operators already use behavioural research and passenger insight of this kind. A Smart Data scheme should build on that capability and enable relevant insights to be used more consistently across operators. This should be supported by appropriate privacy, data-minimisation and equality safeguards, with aggregated data used wherever individual-level data is unnecessary.

The service must distinguish observed conditions from predictions, communicate uncertainty and avoid creating false confidence.

Rail reform is intended to bring infrastructure and passenger operations together under Great British Railways and enable more whole-system decisions across track and train. This should reduce some organisational barriers to sharing operational and passenger data, although Great British Railways is still being established and the benefits will depend on implementation.

Better data will not improve outcomes if the rail system can't act on it. Demand data may reveal persistent crowding, for example, while rolling-stock availability, contractual arrangements and operating constraints limit changes to train formation or capacity allocation. Rail reform and future operating and contractual arrangements should preserve clear mechanisms to review and adjust services as demand changes.

¹ The Department's call for evidence sets out a series of potential Smart Data use cases for respondents to assess. This response refers to those use cases by name. Their full descriptions can be found in the [Smart Data: multi-sector call for evidence](#).

Rail use case 2: Supporting more proactive management of rail infrastructure

The evidence gathered by the Academy for this response strongly supports the objective of moving from reactive or periodic maintenance towards more risk-based, predictive and proactive asset management. The rail sector already collects substantial asset, inspection, condition, maintenance, failure, performance and environmental data and increasingly uses it to support proactive asset management. But remaining fragmentation across systems and organisations, including inconsistent identifiers, can still limit how readily data is combined and reused.

Important progress is already under way. Train-borne measurement, track imagery, remote condition monitoring and predictive tools are already used to identify faults and inform short- and long-term maintenance and investment decisions, and there are examples of train operators and infrastructure managers working together to share live asset condition data. The role of a Smart Data intervention should be to build on and scale this practice by addressing remaining interoperability, access and governance gaps, rather than treating proactive asset management as a new activity.

[Reviving Our Ageing Infrastructure](#), a report from the Academy-led National Engineering Policy Centre. The report identifies accurate data and digital tools as central to proactive infrastructure management and calls for frameworks to share asset condition and performance data securely across infrastructure networks.

- Combining condition, inspection, maintenance, failure and environmental data could identify risks before failure occurs. For example, the effect of drainage on track, earthworks and flood risk.
- Train-mounted monitoring is already used on some newer fleets to collect track condition data. This could reduce some separate inspections and associated line blockages, while helping identify problems earlier. Data from different monitoring systems, including plain line pattern recognition data and data from axlebox and train-mounted accelerometer or inertial measurement units, could become more valuable when combined.
- A shared view of rail assets that includes the dependencies with construction, utilities and other infrastructure could reveal risks earlier and reduce duplication, as the same corridor is monitored by different organisations.
- Better evidence could target maintenance and renewal budgets, reduce unnecessary possessions and interventions, lower whole-life costs and compensation exposure and improve network resilience and passenger reliability.

In addition to data sharing, these benefits depend on decision support. Infrastructure managers need tools and capability to convert interoperable data into operational, tactical and long-term decisions, while retaining the skilled judgement of those who inspect, maintain and operate safety-critical assets.

What types of existing customer data held by traders could inform or enable these use cases?

The emphasis in the call for evidence on customer data only partially fits the rail use cases provided. For rail use case 1, customer-authorised preferences or accessibility requirements

could help tailor journey options. However, the core inputs are operational and public-facing data, such as crowding and seat availability, stations and platforms, onboard and station facilities, accessibility, service conditions, disruption, connections and alternatives.

Aggregated or appropriately consented behavioural and user-research data can add important context. This can help operators provide information and alternatives that reflect the consequences for different users rather than treating every delay alike. Much of this analysis is already undertaken by transport operators. Any new scheme should enable its appropriate use and sharing, with privacy, data-minimisation and equality safeguards.

Rail use case 2 is predominantly an operational and infrastructure data use case. Relevant data includes asset registers and location, inspection and condition records, maintenance and renewal history, faults and failures, work orders, service performance and disruption, associated costs, forward- and rear-facing camera feeds and monitoring such as geotechnical sensors or satellite data. These need to be combined with weather, flooding, catchment, drainage, utilities, telecommunications and adjacent land data.

Infrastructure managers, operators, manufacturers, maintainers, consultants and subcontractors each hold parts of this picture. Raw monitoring data can remain with suppliers while asset owners receive static reports, while construction phase monitoring is not always transferred into operation. The first step should be to map data ownership and responsibility for each use case, including whether the statutory Smart Data powers apply or whether other data sharing mechanisms are required.

Procurement is an important lever. Contracts for rolling stock, monitoring equipment, digital platforms and outsourced services should set clear expectations for data access, interoperability, provenance, reuse, retention and handover, including machine-readable export or application programming interface requirements where appropriate. This would reduce the risk that useful operational data remains with a supplier or becomes costly to retrieve when a contract ends.

What barriers to implementation and possible risks are there to each rail use case?

Rail use case 1: Supporting improved journey planning information across the rail system

Journey information loses value quickly during disruption. Different update cycles and incomplete information about platforms, facilities and accessibility can lead passengers to choose unsuitable routes. Services should distinguish confirmed information from forecasts, show when information was last updated and communicate uncertainty clearly. This is especially important where a passenger relies on a connection or accessible facility.

Rail use case 2: Supporting more proactive management of rail infrastructure

- Rail asset data are spread across legacy systems that use inconsistent identifiers, spatial references, terminology and formats. These differences can prevent datasets from being used together, even when each dataset is accessible.
- Incomplete, obsolete or inconsistent inspection and asset records can make predictive outputs misleading. Data used for safety-critical decisions need clear provenance, timestamps and configuration control, together with information about their quality, limitations and level of confidence.
- Records should preserve the history and rationale of interventions, such as repairs. Change logs should record what was changed, when, by whom, why and under what

conditions.

- Data showing spare capacity in the timetable doesn't establish that the railway can safely run additional services. Capacity decisions must also account for the condition and safe operation of track, signalling and rolling stock, with professional oversight retained.
- Organisations may be reluctant to share asset-condition data that could reveal underinvestment or liability exposure. Cybersecurity concerns and unclear responsibility for inaccurate data or missed interventions could also deter participation.
- Platforms used to combine rail data should work regardless of the system or data owner. Otherwise, it could entrench incumbent suppliers, while high access, accreditation or assurance costs could exclude SMEs and researchers.

How could a Smart Data scheme complement existing data-sharing initiatives in the rail sector? What additional value could it provide, and are there risks that it could duplicate or hinder existing initiatives?

A scheme should complement or, where appropriate, extend existing infrastructure rather than create another repository or require duplicate feeds. The Rail Data Marketplace already provides a discovery and access route, while services such as RealTime Trains and Open Train Times show that data from several sources can be combined for passenger information.

For use case 2, the strongest additional value would be to make infrastructure data usable together. That means a shared infrastructure data framework, with common or compatible asset identifiers and spatial references; minimum metadata and provenance; quality and update requirements; interoperable application programming interfaces; clear access rules; conformance testing. Existing asset management and monitoring systems should remain in place where they work, exchanging selected data through standard interfaces.

Rail data already sit within a complex landscape of discovery services, asset management and monitoring systems and wider transport and infrastructure data programmes led by different organisations. Without clear coordination, a new Smart Data scheme could add another set of standards, interfaces and reporting requirements rather than make the existing landscape more coherent.

Government should therefore define the remaining service gap before creating a new mechanism. This requires answering questions such as which decision can't currently be made, which integration is repeatedly rebuilt, what data can't be found or used, and what measurable improvement a scheme would produce. It should also identify which existing components can be reused, who is responsible for each dependency and where responsibility passes between programmes and organisations. Creating another technically compliant platform without resolving interoperability, quality and accountability would add cost while leaving the underlying problem intact.

Smart Data in multimodal transport

What opportunities and benefits would each use case bring to the relevant sectors? For whom, and why?

The outcomes sought are worthwhile, but the evidence from innovators in the Academy's network doesn't support treating all four use cases as gaps that require a new scheme. For multimodal transport use cases 1, 2 and 3, important data and governance arrangements already exist, and the case for intervention depends on identifying what remains unavailable, unusable or poorly coordinated.

Multimodal transport use case 1: Supporting door-to-door journey planning across modes

Multimodal journey planners already combine conventional public transport data, supported by services such as Traveline and the Bus Open Data Service. Incremental value could come from more reliable real-time disruption and connection information, integration of modes not yet covered, and the careful use of passenger preferences or accessibility requirements. The clearest unresolved public-interest gap identified in our evidence is comprehensive information on the accessibility of bus stops, bus and rail stations, and safe walking or wheeling routes to and from them.

Multimodal transport use case 2: Supporting more evidence-based local transport planning

In England, local authorities can access bus timetable, fares and real-time location data through the Bus Open Data Service and receive aggregate patronage data, with more detailed patronage data may be shared under local agreements. Rail data is also available. Our evidence shows that use of existing data varies significantly between authorities. It's important to distinguish gaps in data from gaps in analytical or delivery capacity. An evidenced data gap is limited or commercially sensitive data from taxis, private hire, shared cycles and scooters. Addressing this gap could help authorities understand unmet demand and improve access to employment, education and healthcare, but additional data won't by itself create stronger local planning capability.

Multimodal data use case 3: Enabling integrated ticketing

Simpler, best-value payment systems can reduce uncertainty and make public transport easier to use. However, multi-operator and multimodal ticketing already operates under established competition law arrangements and local scheme agreements, and evidence points to planned contactless multi-operator capping beginning with bus and light rail. Government should allow this implementation to demonstrate its effect and identify any remaining heavy-rail or last-mile gap before creating a separate Smart Data intervention.

Multimodal data use case 4: Supporting reliability and availability of Electric Vehicle (EV) charging

Making depot charging available to third-party commercial and public-service fleets can allow SMEs to decarbonise vehicles without investing in their own charging infrastructure. Better availability and demand information could improve use of that capacity. However, a fleet operator must prioritise its own vehicles, so mismatches between third-party demand and operational availability can't be solved by data alone. Policy should also avoid improving private car convenience in ways that undermine the wider objective of shifting journeys to public transport or active travel.

What types of existing customer data held by traders could inform or enable these use cases?

Customer data is not the principal input for multimodal transport use cases 1, 2 and 3. Door-to-door planning could use passenger-authorised preferences and accessibility requirements, but the core inputs are timetable, route, fare, vehicle-location, disruption, interchange, facility and accessibility data. Local transport planning requires aggregated service and patronage data, supplemented, where justified and appropriately governed, by data from connected and

shared mobility providers. Integrated ticketing depends primarily on standardised fare, ticketing and operator data within existing schemes.

Aggregated demand data, such as ticketing and booking records, passenger counts, origin-destination and time-of-day patterns, journey-planning searches and demand-responsive transport requests can help authorities plan services and identify changing demand. Analysis should distinguish observed use from unmet or suppressed demand, which may not appear in existing service data.

For multimodal transport use case 4, relevant vehicle and battery data could tailor charging options, while charge point availability, pricing and waiting-time data describe the service. Our evidence also highlights the need for aggregated data about expected take-up by commercial and public service fleets so infrastructure providers can assess demand. Data collection should be limited to what the defined service needs and shouldn't substitute for investment in physical capacity.

What barriers to implementation and possible risks are there to each multimodal use case?

Multimodal transport use case 1: the main risks are incomplete modal coverage, inconsistent real-time updates and accessibility data that is technically available but not reliable enough to support an end-to-end journey. Digital-only services can also exclude some passengers, so analogue options should remain an option.

Multimodal transport use case 2: commercial sensitivity and lack of consistent data from taxis, private hire and shared mobility can leave gaps. Local analytical and delivery capability varies, so a data mandate without support to use the data may have limited impact.

Multimodal transport use case 3: a new scheme could duplicate or disrupt ticketing arrangements already governed through the Public Transport Ticketing Schemes Block Exemption and local agreements, particularly while contactless capping is being implemented.

Multimodal transport use case 4: in the fleet-owned depot model described in our evidence, the operator faces real capacity and scheduling constraints and must prioritise its own vehicles subject to contracts. There is also a strategic risk of incentivising private car use at the expense of public transport.

Across all four use cases, government should distinguish missing data from missing incentives, analytical capability, physical infrastructure or organisational coordination. Smart Data can address the first category and sometimes the second, but it can't by itself resolve the others.

Data-driven and AI-enabled services can reproduce bias if historic service patterns or digital datasets under-represent disabled, older, low-income, rural or digitally excluded passengers. Scheme design should embed measures to address that from the outset and monitor whose journeys remain missing from the data.

How could a Smart Data scheme complement existing data-sharing initiatives in multimodal transport? What additional value could it provide, and are there risks that it could duplicate or hinder existing initiatives?

Government should begin by mapping existing mandates, feeds, platforms and scheme agreements. These include the Bus Open Data Service, rail-data services, multimodal journey planners, local data sharing agreements and the established framework for integrated

ticketing. The question should be what users still can't do, rather than whether another data platform can be created.

A Smart Data intervention could add value by filling clearly evidenced gaps, such as bringing additional modes of transport into compatible feeds; establishing comprehensive and usable stop, station and safe access route information; enabling consented use of accessibility preferences; providing the common back-office interfaces needed where rail joins integrated ticketing. For multimodal transport use case 2, it may also establish purpose-limited access to commercially sensitive mobility data where existing agreements can't scale.

Existing systems should remain the delivery route where they already work. New requirements should extend compatible standards and governance rather than force participants to maintain parallel datasets or displace schemes that are close to implementation.

What governance-related barriers or enablers are most significant in practice for enabling data sharing in the transport sector?

Existing arrangements already govern much of the data sharing needed for the multimodal use cases. In England, the Bus Open Data Service requires bus operators to supply timetable, fares and real-time location data through a standard route, enabling its reuse by journey planning services. Local authorities already receive aggregate bus patronage data and can obtain more detailed information through data sharing agreements. Integrated ticketing operates under competition law rules and local scheme agreements, which set out how data may be used and responsibilities allocated.

For EV charging, our evidence identifies an operational constraint rather than a data governance gap – where a charging provider also owns a fleet, this risks being prioritised subject to contractual agreements.

- Any new intervention should address a demonstrated gap and provide a clear operational or passenger benefit. A parallel data feed or additional approval process would add costs and could deter participation.
- Fragmented ownership, inconsistent standards and unclear access arrangements are significant barriers. Common requirements for APIs, metadata, access and security could apply across transport, while rail-specific definitions and safety requirements should remain specific to the sector.
- Shared data should include information about its source, timestamp, quality and limitations. Data that can be accessed but can't be interpreted or relied upon won't support effective or safe operational decisions.
- Approval and assurance requirements should reflect how data will be used. Read-only access shouldn't face the same requirements as a service that informs maintenance or safety-critical decisions. Excessive fees or assurance requirements could exclude SMEs and researchers.
- Commercially sensitive mobility and asset condition data don't need to be made openly available. Existing local and ticketing agreements show that data can instead be shared with agreed users for specified purposes, subject to appropriate access and security controls.

How should liability and accountability be managed where transport data are shared between multiple providers and users, and responsibilities are shared across different actors?

Liability should follow the function performed, with a traceable chain from source data to action:

- Data generator or provider should be responsible for recording and supplying data accurately, with provenance, timestamps, quality status and stated limitations.
- Third-party service should be responsible for transformations, integration, models, presentation and recommendations, including how it handles uncertainty or conflicting sources.
- Operational or infrastructure duty holder should be responsible for the final operational and safety-critical decision. Rail reform or a new data scheme should not blur the established regulatory oversight of these duties.

Existing transport arrangements offer practical models: scheme agreements can specify data uses and responsibilities, and multi-party data-sharing agreements can permit onward sharing for defined purposes. These should be standardised where repeated, while safety cases and rail-specific liabilities remain with the appropriate duty holders.

Governance should avoid layering multiple overlapping approval and assurance processes onto the same service. A shared model for common functions, such as participant accreditation, baseline security, standard contractual terms, audit evidence and incident protocols, could reduce duplication and give organisations stronger incentives to cooperate. Requirements should remain proportionate to the risk and use of the data, without weakening rail-specific safety cases.

How could Smart Data support or enable use cases in other modes of transport not referenced in the call for evidence? What are the key opportunities and barriers?

Taxis, private-hire vehicles, shared cycles and scooters could improve first- and last-mile options and give local authorities a more complete view of travel. Their strongest contribution may be to accessible end-to-end journeys, for example, where a bus or rail service can't accommodate a passenger's needs.

Government should first make the best use of reliable bus and rail data, then add transport modes where the user benefit is clear and sufficiently accurate data are available. Barriers include commercial sensitivity, inconsistent coverage and standards, uncertain accessibility, and the risk that an elaborate integration adds little value to services users already access on demand.

Agrifood supply chains

What data sets do you use as part of your role? What barriers do you face when accessing, using or sharing this data?

Our evidence covers data from primary production through storage, processing, procurement, consumption, waste and nutrient recovery. Relevant datasets include crop, soil, weather and farm management data; sensor, drone and satellite observations; yield and quality measurements; site, storage-unit, commodity, lot and inventory records; temperature, moisture and carbon dioxide measurements in stored grain; samples and laboratory results;

treatment, drying and aeration records; procurement, pricing, logistics and traceability data; food waste, environmental and nutrient information.

This data is distributed and often created through activities involving several organisations, so no single organisation necessarily holds the complete record or has sole control over its use. Farmers and storage operators provide operational context; equipment and platform providers hold sensor data; laboratories and assurance bodies hold test results; logistics providers hold movement data; processors and buyers hold acceptance and quality information; government systems hold land, payment and environmental records. Practical control therefore depends not only on ownership or entitlement, but on usable export or application programming interfaces, permissions, identifiers and metadata.

Our evidence points at manual work to discover, obtain, reconcile and re-enter data that already exist. Barriers include proprietary dashboards, spreadsheets and PDFs; fragmented identifiers and definitions; poor links between public and private data; intermittent rural connectivity; limited SME technical capacity; uncertainty about rights in raw data, processed data, models and outputs. The challenge is access and interoperability, not simply data availability.

Which parts of the agrifood sector should be prioritised for Smart Data intervention, and why?

Priority should be given to primary production, the interface between production and processing, and post-harvest storage and quality assurance. This is where important traceability, sustainability, forecasting and quality data originate; where data are especially heterogeneous; where delayed or poorly informed decisions can lead to physical loss, quality downgrades, safety risk, avoidable treatment and energy use or rejected produce.

Government should also test applications in local procurement, food-waste management and circular nutrient recovery. These could connect local demand with production planning and link information on waste provenance and composition to soil, crop and nutrient needs. These broader use cases are valuable as they cross organisational and sector boundaries, but they should follow the same test of legal fit, additionality and benefits for intended users and the organisations asked to provide data.

Prioritisation should also consider how costs and benefits are distributed. A use case is unlikely to be sustainable if farmers or smaller operators bear most of the cost of collecting and maintaining data while larger downstream organisations capture most of the value.

What specific data access challenges exist in the agrifood sector and what action could address them? How might these challenges affect a Smart Data scheme?

The main access challenges include finding what data exist; establishing entitlement and permission; obtaining machine-readable data; linking government, farm and supply chain systems; matching records across organisations; operating where connectivity is intermittent. A dashboard view or one-off PDF doesn't amount to effective portability. Nor should the scheme assume that every participant will have the connectivity, infrastructure or technical capacity to use continuously connected cloud services.

Government should:

- map the data, holders, rights and access routes needed for each priority use case before setting scheme requirements

- establish purpose-specific, revocable and auditable permissions, including clear rules for onward sharing and retention
- require usable application programming interfaces and bulk or local export options, while supporting offline collection and synchronisation
- develop common identifiers and practical tools that allow legacy and smaller systems to exchange data without being replaced
- test the legal applicability of each use case early. Much operational agrifood data is created through activities involving several organisations and may not arise through a conventional trader-customer relationship. Where the statutory powers don't fit, voluntary, contractual or trusted-data-space mechanisms may be needed.

What specific data quality, accuracy or completeness issues exist in the agrifood sector today? How might these affect a Smart Data scheme?

The most important gap is often consistent meaning, identifiers and provenance. A soil moisture value, for example, is of limited use without the sensor type, installation depth, calibration, location, timestamp and quality control. A grain measurement can be misleading if a sparse sample is treated as representative of an entire bulk. Similar problems arise where laboratory results can't be linked to the relevant sample, lot and storage history.

A scheme should distinguish availability, interoperability, provenance and fitness for purpose. It should also distinguish directly measured, estimated or modelled, calculated or derived, manually entered and laboratory-verified data. Minimum metadata should include identifiers, units, timestamps, measurement or sampling method, calibration status, spatial and temporal resolution, processing history, stated uncertainty, quality flags and version history.

Quality requirements should be proportionate to the decision. A record used for general benchmarking doesn't require the same assurance as data driving irrigation, storage treatment, food-safety or finance decisions.

What use cases could a Smart Data scheme enable in agrifood, what outcomes could they deliver, and what barriers and governance structures would need to be addressed?

The Academy's evidence identifies five agrifood use cases. They should be treated as use cases to test, rather than assuming that a statutory Smart Data scheme is necessarily the right delivery mechanism.

Use case	Data and service	Potential outcome
Reusable production record	A permissioned record for a field, crop or production cycle, linking location, variety, dates, environmental conditions, management, yield and quality.	Capture core information once and reuse it selectively for farm management, research, assurance, lending or reporting.
Production history to crop quality	Link quality measurements at harvest or processing with variety, field conditions, environmental history and management; return	Reduce rejected produce, improve pricing and processing decisions, and

	processor requirements and quality feedback to producers.	give producers practical value in return for data.
Grain condition and intervention management	Combine temperature, moisture and carbon-dioxide readings with lot, storage asset, weather, equipment, inspection, sampling and treatment history.	Detect deterioration earlier, target aeration, drying or treatment, reduce loss, labour and energy, and retain evidence for decisions.
Portable quality and chain of custody	Connect sample location and method, condition data, laboratory results, certificates, storage and treatment history, and subsequent movement and custody.	Allow buyers, processors, laboratories, assurance bodies, lenders, insurers and regulators to reuse trusted evidence rather than re-enter or reverify it.
Local procurement and circular nutrient recovery	Connect demand and procurement with production planning, logistics, consumption and waste. Link waste provenance and composition with soil, crop and nutrient needs.	Support local food resilience, reduce waste, improve nutrient recovery and align production more closely with demand.

The above use cases depend on common foundations, such as reliable identity and permissions; shared identifiers and semantics; provenance and quality information; secure interfaces and export; and clear allocation of responsibility between source, service and decision-maker. Agrifood expertise is needed to define sector-specific standards for fields, crops, lots, storage assets, samples, measurement methods, treatments, certificates, traceability, food safety and environmental performance.

Governance should include farmers, SMEs, technology providers, processors, laboratories, inspection and assurance bodies, logistics operators, researchers and relevant public bodies. A scheme dominated by the largest platforms or downstream buyers would be unlikely to earn trust. Use cases involving finance or insurance should allow trusted derived indicators where these meet the purpose, rather than routinely exposing the full underlying dataset.

To what extent do commercial arrangements constrain data sharing, and what should a Smart Data scheme include to address these constraints?

Commercial constraints are substantial. Stock levels, crop or grain condition, expected losses, quality problems, treatment history, pricing and supplier relationships can affect negotiations and liability. This data is also often created through activities involving several organisations, while contracts and platform terms may leave rights in raw measurements, processed datasets, derived outputs and models unclear. Vendor lock-in and costly one-off work to connect systems compound these issues.

A producer or smaller operator may be willing to share data for an advisory, quality or storage service, but not for pricing, benchmarking, marketing, credit, insurance or training a proprietary model. Permission for the immediate service should therefore be specific, minimal,

time-limited where appropriate, revocable and recorded. Secondary use, including AI model training, should require a separate explicit agreement.

The scheme should make customer-generated operational data portable without requiring providers to disclose algorithms, source code or device designs. It should also address reciprocity: where farmers and small operators bear the cost of generating and maintaining data, they should receive practical benefit such as better advice, quality feedback, reduced reporting, market access or more appropriate services. Tiered access and trusted derived results can reduce commercial exposure while enabling a defined decision.

How could a Smart Data scheme complement existing agrifood data-sharing and digitalisation initiatives? What value could it add, and could it duplicate or hinder existing work?

A Smart Data scheme shouldn't create another central agrifood platform. Useful government systems, farm management tools, sensor platforms, satellite and weather archives, laboratory systems, assurance arrangements and research repositories already exist. The Rural Payments Agency land interface also demonstrates that permissioned, machine-readable access to agricultural information is technically feasible.

The additional value should be a common trust and interoperability layer across existing systems: standard definitions and identifiers; secure authentication and authorisation; purpose-specific permissions; minimum interface and export requirements; quality and provenance metadata; conformance testing; accountability and redress; and adapters for legacy systems. This should complement the Food Data Transparency Partnership and Food Data Infrastructure and Governance work.

Success should be measured by whether users can discover, access, combine and act on data more quickly; whether repeated reporting and costly one-off work to connect systems fall; whether loss, energy use or rejected produce decline; and whether smaller providers can offer services. The number of new datasets placed on a platform wouldn't by itself demonstrate value.

Cross-sector governance

What aspects of governance are most important for Smart Data schemes in practice, and why?

We support a federated approach. Common requirements should be coordinated across Smart Data schemes, while sector bodies and regulators remain responsible for sector-specific data definitions, quality and safety. Central coordination should focus on common standards and shared functions rather than requiring every sector to adopt a new platform.

In practice, governance should address the following:

- Before designing a scheme, government should identify the problem it is intended to solve, the data required, the legal basis for sharing them and the gap in existing arrangements. It should also specify who should benefit and how success will be judged. This is important because some valuable operational and infrastructure data identified by contributors may not arise from a trader-customer relationship, while other data is already governed through statutory requirements or existing agreements.
- Requirements for identity, authentication, permissions, audit trails, cybersecurity, incident management and basic interfaces should align across schemes. Reusing these

functions would make cross-sector services easier to operate and avoid each scheme developing a separate trust framework.

- Sector bodies and regulators should define what sector data mean, what level of quality is required and who is responsible for keeping them current. Rail needs its own asset and safety standards; transport services need consistent definitions for fares, accessibility and service information; and agrifood needs standards covering matters such as lots, traceability, measurement methods, calibration and uncertainty.
- Governance should preserve a traceable record from the original data through any processing or analysis to its final use. The organisation supplying data should be responsible for recording and providing them accurately, third-party services should be accountable for their processing and recommendations, and established duty holders should remain responsible for final operational and safety-critical decisions.
- Permissions should be limited to a stated purpose and the minimum data required. They should be revocable and recorded in an audit trail. Approval and assurance requirements should reflect the risk of the proposed use, so that read-only access does not face the same requirements as a service informing maintenance or safety-critical decisions.
- Data generators, SMEs, technology providers and relevant sector experts should be represented in standard-setting and governance. Arrangements dominated by the largest data holders or buyers would be less likely to command trust and could concentrate market power.

How should Smart Data schemes be governed and coordinated in practice where services or data flows span more than one sector?

Coordination is needed at two levels: around each cross-sector service and between the government data programmes on which those services depend.

For each cross-sector service, one organisation should be identified to coordinate delivery. Participating schemes, regulators and organisations should agree how data will move between them, the legal basis and permitted purposes for sharing it, where responsibility passes between organisations, and how incidents will be handled. The coordinating organisation should oversee the service as a whole without displacing existing regulatory or operational responsibilities.

Government should convene the organisations required for each use case, clarify how different legal and regulatory frameworks interact and resolve gaps or conflicts between schemes. Sector regulators and implementation bodies should agree how they will work together at the boundaries of their remits. Industry and public bodies should define the operational requirements, implement the necessary interfaces and maintain the data they contribute.

Some cross-sector services will depend on Smart Data alongside public sector and other relevant infrastructure data programmes. Government should use a joint delivery framework to clarify what each initiative will provide, which components can be reused, where responsibility passes between them and how shared costs will be met.

An emerging example shared through the Academy's engagement concerned a proposal in Liverpool City Region to link NHS outpatient appointment data with DWP eligibility data so that eligible patients could automatically receive an electronic bus ticket. Although this is not

a conventional Smart Data scheme, it illustrates the potential value of coordinating data and delivery across health, welfare and transport. It also shows the need for clear arrangements covering the legal basis for sharing data, responsibility for determining eligibility, delivery of the service and accountability when organisations span different sectors.

What factors would encourage and deter organisations from participating in a Smart Data scheme?

Participation is most likely where the value includes better operational decisions, fewer repeated inspections or reports, reduced loss or energy use, easier integration, access to new customers, or improved service quality. Clear rights to retrieve and move operational data, predictable technical requirements and liability, and the ability to grant narrow permissions would also encourage participation.

Deterrents include costly one-off work to connect systems; unclear ownership, liability or permitted use; accreditation and reporting burdens; cybersecurity exposure; lack of local connectivity or capability; loss of intellectual property; and concern that commercially sensitive data will be used for a different purpose. Participation will also be weak where the cost of producing and maintaining data falls on a farmer, operator or small supplier while value accrues mainly to a large buyer, platform, lender or insurer.

Governance should:

- tier accreditation and assurance by role and risk, for example, distinguishing read-only access from analytics or safety-critical decision support
- provide low-cost standard interfaces and practical tools that allow older or smaller systems to exchange data, including local or offline export options where continuous connectivity is unavailable
- make fees and compliance requirements proportionate, with practical participation routes for SMEs, researchers and new entrants
- protect purpose limitation and require separate permission for materially different secondary uses
- preserve legitimate provider intellectual property while making user-generated operational data portable
- include smaller data generators and service providers in standards and governance, not only the largest data holders.

What lessons from existing UK data-sharing initiatives should shape Smart Data governance?

- The Bus Open Data Service shows that a clear mandate and standard route for supplying data can support reuse by journey-planning services. It makes bus timetable, fares and real-time location data available in England. However, the evidence received noted that use of these data varies significantly between local authorities. Making data available does not ensure that it will be used consistently.
- The Rail Data Marketplace provides a useful way to discover and access data, but rail contributors still encounter infrastructure data spread across legacy systems and suppliers, inconsistent identifiers, and static reports that cannot readily be combined.

Future schemes should focus on making data usable together, rather than only making individual datasets easier to find.

- Public-transport ticketing shows that competition-law rules and scheme agreements can define permitted data uses, revenue allocation and responsibilities. Local multi-party agreements can also allow data to be shared for specified purposes. Government should build on these arrangements and avoid introducing overlapping requirements while contactless multi-operator and multimodal ticketing systems are still being implemented.
- The Rural Payments Agency land API shows that permissioned, machine-readable access to agricultural data is technically feasible. However, access to one interface does not resolve difficulties with discovering data, securing permissions, reconciling identifiers and metadata, or connecting public and private systems. Smart Data should establish common standards and permissions that connect existing systems, while complementing the Food Data Transparency Partnership and Food Data Infrastructure and Governance work.
- Open Banking shows the value of common rules, trusted standards and access through authorised third parties. These principles can inform other Smart Data schemes, but data definitions, quality requirements and assurance arrangements must reflect each sector. The evidence received shows that operational agrifood data and safety-critical rail data raise different questions about control, risk and responsibility.

Wider considerations

Are there any other risks, opportunities or benefits, including environmental or societal impacts, that relate to Smart Data?

In rail, better integrated asset and environmental data could support climate resilience, earlier intervention, fewer possessions, lower embodied carbon and reduced personnel exposure. In transport more broadly, clearer information and payment can improve access to employment, education and healthcare, passenger confidence and public-transport uptake. In agrifood, trusted data could reduce crop and grain loss, energy and treatment use, repeated reporting and food waste, while strengthening traceability, local procurement, food system resilience and circular nutrient use.

The corresponding risks should be designed into evaluation from the outset:

- poor quality, delayed or decontextualised data can create unsafe decisions or false confidence
- digital-only or cloud-only participation can exclude people, farms and businesses with limited connectivity, equipment or capability
- function creep can turn data shared for an operational service into commercial surveillance, discriminatory pricing, credit or insurance decisions, or model training without meaningful permission

- dominant platforms or downstream buyers can gain bargaining power if interoperability and portability are weak
- over-specified technical standards and overlapping assurance processes can lock schemes into early design choices, exclude smaller participants and impose cost without reducing material risk
- better data can expose a clear need for change while contracts, funding, fleet availability or operating rules leave organisations unable to act. Schemes should be paired with proportionate mechanisms to review and adapt services
- new platforms and assurance processes can entrench incumbents or impose costs without improving outcomes
- cybersecurity failures can interrupt services or expose operationally sensitive information
- optimising one outcome in isolation can undermine another - for example, making private EV journeys easier while weakening modal shift to public transport.

Government should evaluate schemes against a set of operational, economic, environmental, safety, accessibility, competition and inclusion outcomes, including whether organisations can act on the insight produced, and publish limitations and unintended effects as well as benefits.