Good practice guide

# Improving government data:
## A guide for senior leaders

National Audit Office

July 2022

Design & Production
by Communications Team
DP Ref: 012207

We are the UK's independent public spending watchdog

# **Improving government data:** a summary guide for senior leaders

**Who is it for?**

This guide is for senior leaders responsible for delivering government services.

**Aim of this guide**

Our aim is to encourage decision-makers to realise the benefits of better use of data by helping them understand in more detail the core issues to be addressed which have held back progress in the past.

**Where it applies**

We focus on data to support the operational delivery of public services, but much of our guide will also be relevant to data for decision-making and to improve performance.

## Further information

If you are interested in the NAO's work and support for Parliament more widely, please contact:

Parliament@nao.org.uk
020 7798 7665

## The problem areas

Our guide discusses overcoming barriers in the following areas:

### Data-sharing

**Why can't departments just put in data-sharing agreements?** Agreements must consider all the practical considerations discussed in this guide, including quality, technical limitations and the cost and effort the providing organisations incur to make the data available.

### Data quality

**Why does data quality matter?** Data cannot be trusted otherwise. Data collected by one part of government may not be of sufficient quality to be used by a different part of government for a different purpose. Government's Data Quality Framework offers a more structured approach to improving the quality of data held by departments.

**Why does the data quality problem persist?** During the pandemic we saw examples of non-personal data being aggregated for beneficial outcomes, for example on the number of people sleeping rough or at risk of doing so. Personal data presents more challenge, especially within the limits of existing systems. Activities delivering tactical solutions can make it harder to achieve the overall strategic aims.

### Data standards

**Why are standards so hard to implement?** The structure of government is heavily siloed and departments have a high degree of autonomy. Legacy systems make it difficult to introduce standards into this environment and government has struggled to make substantial progress over the past 20 or so years.

### Resourcing

**Why can't we put more resources into tackling the problem?** Departments rarely measure inefficiencies in terms of the time and cost of improving poor-quality data. It can be difficult to make the case for funding stand-alone data projects. Current government funding and performance-monitoring arrangements do not support cross-departmental working, especially where one or more departments need to contribute resources to support another department to achieve its objectives.

### Access to raw data and APIs (application programming interfaces)

**Why can't departments just give each other access to raw data or via APIs?** Without further information about that data, users may not understand key contextual factors relating to quality and limitations in collection. People may misuse the data leading to poor outcomes and reputational damage.

### Creating cross-government data sets for multiple users

**Why can't government create single data sets from each other's data?** Merging personal data which does not easily match is difficult. Further questions arise around ownership, maintenance, funding, privacy, and the risks arising from data aggregation.

### Data analytics

**Why can't data analytics solve the problem?** Data analytics and tools work well with good-quality data although effort is required to engineer the data when it comes from disparate sources. But there are situations where the accuracy and integrity of the data will make analytics difficult to apply, especially for personal data.

## The way forward

### Reviewing why initiatives have failed in the past

To move forward government needs to acknowledge and address the difficulties we highlight in this guide and understand why previous initiatives have not succeeded.

### Embedding data standards

A look at the costs and time spent on working around discrepancies in systems should provide insight on how and where to focus resources. Once standards have been established, adoption requires a carefully considered plan, or government risks failing to take the right steps towards interoperability in the short-to-medium term and making it harder to achieve in the future.

### Improving data quality

A variety of frameworks and maturity models are available for organisations wishing to improve the quality and use of their data. This guide signposts them.

### Addressing legacy issues

Government must identify what it has under management across the sector. Departments should identify those main ageing IT systems that, if fixed, would allow government to use data better. Next, they should ensure that modification or replacement of these systems is done with full consideration of how they will support better use of their data.

### Enabling data-sharing

The Open Data Institute has recently published Assessing risks when sharing data: a guide to help organisations identify data-sharing issues and undertake related risk assessments. Where the results indicate that data is not suitable for sharing, there is value in explicitly articulating the reasons. This can help understand the barriers to re-using data and what can be done to address them.

## Conclusion and key messages

- Government data is a leading cause of inefficiencies.
- Organisations need to acknowledge the importance of fixing underlying data issues.
- Data needs to be managed and this cannot be achieved without focused effort, funding and prioritisation.
- Initiatives tend to peter out when the going gets tough, but it is important to continue.
- Fully addressing the issues highlighted in this guide could bring substantial benefits by enabling reform and transformation.

# Introduction

This guide is aimed at accounting officers, chief executives, director generals, directors and chief operating officers and people responsible for government services.

Government wants more effective use of data and data-sharing across public services. Data is government's biggest asset and it is critically important to improve data in order to be able to share that data better and exploit the benefits, which include:

- improved service quality;
- reduction in the burden for citizens;
- greater efficiency;
- reduced costs of public services; and
- improved ability to measure the impact of policies and programmes.

In 2019, the National Audit Office (NAO) published a report on *Challenges in using data across government*.[1] We said that getting the right data in the right place at the right time is a basic driver of value for money in government: making services work for people, improving government's systems and processes, and supporting better decisions.

But for established organisations, these benefits are not simple to achieve in practice. Senior stakeholders in many organisations in both private and public sectors are struggling to understand the complexity of the existing data challenge and the extent of the investment needed. It requires money, time and skills to fix the problems and exploit data assets. Senior decision-makers may sign off multi-million-pound investments in technology without realising that they are most likely compounding the data problem. New systems may add to existing issues if insufficient thought is given to how data can be exchanged in a fragmented data landscape. Data professionals who may be in a position to address the challenges can find themselves reporting into an organisation which may not be listening to them.

Organisations that understand and have succeeded in overcoming the data challenge fall into one of two broad categories. Firstly, there are those which are designed and built for data exploitation from the outset and do not carry the 'baggage' of legacy systems and ways of working. Examples include Google, Amazon and Netflix. As a result they are naturally able to exploit their data assets and can readily take advantage of business intelligence, advanced analytics and artificial intelligence. Secondly, there are organisations with legacy systems which have been forced to address the data challenge in response to external events. For example, following the financial collapse of 2008, the financial services sector was subject to additional regulatory obligations. While these focused predominantly on data to support risk models, many financial services firms began to see substantial benefit in being able to exploit their data for other purposes and this has been a significant enabler of digital transformation in those organisations.

We are entering a phase where most organisations now recognise the benefits of data exploitation but do not have the modern data architecture to support it. Rather, they have complex, legacy data architectures. They also do not have a compulsory purpose such as regulatory compliance to justify the investment needed. A useful indicator of how seriously organisations take the data challenge is their spend on data as a percentage of their overall technology budget, if indeed such spend is separately identified and monitored.

1 Comptroller and Auditor General, *Challenges in using data across government*, Session 2017–2019, HC 2220, National Audit Office, June 2019.

Barriers standing in the way of better use of data are not obvious or simple to overcome. In our report we identified three substantive issues:

- **The quality of data is not well understood.** Quality data is not a 'free good'. While government has repeatedly talked about the benefits arising from better use of data, new initiatives often expose the poor quality of the data itself and the lack of a structured approach to address the underlying causes.
- **Data is not always seen as a priority.** It can be challenging to make the case for long-term investments to improve the quality and sharing of data. A lack of understanding of the costs involved in cleaning, combining and improving data within the constraints of existing systems exacerbates the challenge. People do not monitor the time or costs involved in sorting poor-quality, disorganised data.
- **There is a culture of tolerating and working around data that is not fit-for-purpose.** Government has lacked the necessary capability, leadership and culture to introduce and support sustained improvements.

There is still a poor appreciation of the state of the data in legacy systems and its impact on the transformation of operational services. Data issues include age, quality and consistency across different systems. More than half of all civil servants work in operational delivery – although the cost and inefficiency of their working with poor-quality data has not been quantified, it is likely to represent a significant overall expense.[2] In our publication *Improving operational delivery in government: A good practice guide for senior leaders*, we said that in 2020-21 central government departments alone expected to spend £456 billion on the day-to-day running costs of public services, grants and administration.[3] This highlights the potential for efficiencies to be gained from the improved use of data.

In September 2020, government published its *National Data Strategy*.[4] This sets out a framework for action for government on the role of data in public services, the economy and society more widely. The strategy identifies four areas to be addressed – data foundations, data skills, data availability and responsible data use.[5] It also identifies five areas of action, one of which is to transform government's use of data to drive efficiency and improve public services.[6] The strategy refers to the NAO report on *Challenges in using data across government* illustrating the existing problems and sets out actions for the centre of government and departments to take.

The strategy says government will set out its implementation steps in future updates. The Department for Digital, Culture, Media & Sport (DCMS) has published a progress report in the form of a monitoring and evaluation framework. The Central Digital and Data Office (CDDO), part of Cabinet Office, has published *Transforming for a digital future: 2022 to 2025 roadmap for digital and data* which includes a mission on better data to power decision-making in government.[7] CDDO is also driving several important strands of activity.

This guide is intended to be complementary to the *National Data Strategy*. However, as we show, there have been numerous previous attempts to achieve similar aims but tangible progress has been slow because of the complexity and enormity of the task. The aim of this guide is to encourage decision-makers to realise the benefits of better use of data by helping them understand in more detail the core issues to be addressed which have held back progress in the past. As with our *Challenges in using data across government* report, our focus is on the use of data to support the operational delivery of public services, but many of our observations will also be relevant to data to support decision-making and improve performance.

2 According to Civil Service Statistics 2021, 48.3% of civil servants are reported to work in the operational delivery profession, available at: www.gov.uk/government/statistics/civil-service-statistics-2021, (link accessed 12 July 2022). However, this excludes data from the Department for Work & Pensions, which is the largest department by headcount, thereby understating the true position. The operational delivery profession itself estimates that the actual figure is closer to 70%.

3 National Audit Office, *Improving operational delivery in government: A good practice guide for senior leaders*, March 2021. Available at: www.nao.org.uk/report/improving-operational-delivery-in-government/, (link accessed 12 July 2022).

4 Available at: www.gov.uk/government/publications/uk-national-data-strategy, (link accessed 12 July 2022).

5 These are referred to as 'pillars' in the strategy.

6 These are referred to as 'missions' in the strategy.

7 Available at: www.gov.uk/government/publications/roadmap-for-digital-and-data-2022-to-2025, (link accessed 12 July 2022).

# The problem areas

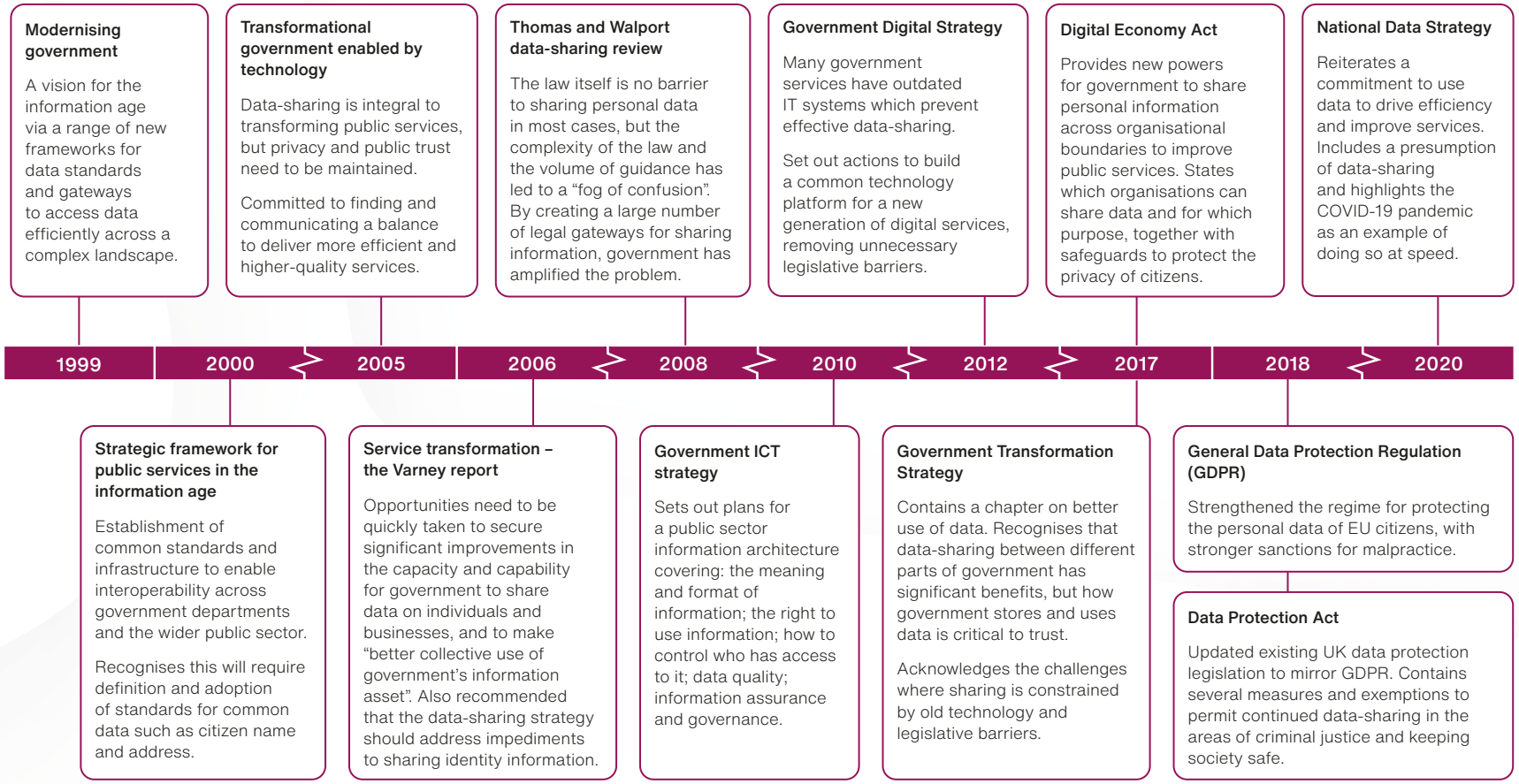
Government knows it can do much more with data, for example greater sharing, personalisation of services, reducing the burden for citizens, analytics and exploiting cutting-edge technology. Benefits could include:

- better delivery of interconnected public services, such as tax, welfare and pensions and health and social care provision;
- reduced burden on citizens through not having to provide the same information repeatedly to different parts of government;
- reduction in inefficiencies that arise from duplicated or poorly coordinated information;
- better evidence to assess whether policies are having their intended effect and interventions are effectively designed; and
- greater ability to solve problems quickly, as was seen during the recent COVID-19 pandemic.

There have been many strategies and initiatives over the past 20 years to promote joining up and using data more efficiently across different services. Initiatives have included proposals for data frameworks and standards, promoting interoperability of systems, and measures to overcome legislative and cultural hurdles. In practice, despite the good intent, tangible progress has been slow as shown by repeating themes in successive strategies (**Figure 1**).

## Figure 1

Government strategies and initiatives on using data across government




Source: National Audit Office analysis of government initiatives to improve data

# Data-sharing

Government's long-held ambition for better data-sharing to achieve efficiencies in public services is supported by the following legislation:

- Digital Economy Act 2017 – provides a basis for data-sharing where no legal gateways exist and consent cannot be relied on or is not appropriate.
- Data Protection Act 2018 – includes a direction to the Information Commissioner's Office to prepare a code of practice for data-sharing (published in 2020).

Important considerations needed when considering data-sharing are:

- consent, ownership, and privacy;
- safety and security;
- use and misuse; and
- data management and integrity.

## Why can't departments just put in data-sharing agreements?

In April 2017, government passed the Digital Economy Act to enable the sharing of personal data between public authorities to improve how public services are delivered for the well-being of individuals and households. Statutory codes of practice which follow data protection principles ensure that sharing personal data is proportionate.[8]

However, in our report *Challenges in using data across government*, we found that the Digital Economy Act had not so far given departments the reassurance they need to be confident about sharing data legally. Departments told us they would welcome more support on how to use the Act appropriately to support data-sharing.

We also frequently hear that departments are grappling with ethical issues and that these also remain a major inhibitor to sharing data in practice.

Data-sharing agreements must take account of all the practical considerations discussed in this guide. These include any quality and technical limitations as well as the cost and effort on the part of the providing organisation to make the data available.

8 Digital Economy Act 2017, available at: www.gov.uk/government/publications/digital-economy-act-2017-part-5-codes-of-practice, (link accessed 12 July 2022).

# Data quality

## Why does data quality matter?

High-quality data is essential for effective service delivery, performance monitoring and improvement. The key problem is **data quality** – without this, data cannot be trusted. In December 2020, government published a Data Quality Framework with a stated aim to provide a more structured approach to understanding, documenting and improving the quality of data held by departments. The framework states that at a high level, data quality can be thought of as 'fitness for purpose' and being good enough for what the user intends to use it for. The framework sets out some core characteristics (known as 'dimensions') of data quality (**Figure 2**).

An important consideration in data being 'fit for purpose' is that data collected by one part of government for its own existing processes may not be of sufficient quality to be used by a different part of government for a different purpose from that envisaged at the point of collection. In our report *Challenges in using data across government*, we cited the example of Real Time Information (RTI) where HM Revenue & Customs (HMRC) shares payroll information received from employers with the Department for Work & Pensions (DWP) so that DWP can calculate entitlement to Universal Credit. Occasionally an employer may make a duplicate RTI submission. This is not time-critical for HMRC as it can sort out the data later. For DWP it affects a claimant's entitlement, so DWP incorporated an additional check into the interface to identify duplicates.

**Figure 2**
Characteristics for measuring data quality

| Dimension | Explanation |
|---|---|
| Completeness | The degree to which records are present and contain important data. |
| Uniqueness | No duplication in records. |
| Consistency | The degree to which values in a data set do not contradict other values relating to the same thing. |
| Timeliness | The data is an accurate reflection of the period represented and the values are up-to-date. |
| Validity | Data is within the expected range and format. |
| Accuracy | The degree to which data matches reality and is free from bias. |

**Notes**

1 These dimensions are based on the paper *Defining Data Quality Dimensions* published by the Data Management Association (DAMA) UK Working group.

2 As the framework itself notes, other organisations may describe quality dimensions differently. See, for example, *Dimensions of Data Quality Research Paper*, DAMA Netherlands, September 2020, which surveyed various dimensions by different sources (available at: www.dama-nl.org/wp-content/uploads/2020/09/DDQ-Dimensions-of-Data-Quality-Research-Paper-version-1.2-d.d.-3-Sept-2020.pdf), link accessed 12 July 2022.

Source: Government Data Quality Framework, December 2020 (available at: www.gov.uk/government/publications/the-government-data-quality-framework/), link accessed 12 July 2022

## Why does the data quality problem persist?

The pandemic highlighted the importance of high-quality data. Some commentators have observed that people were able to achieve more in a matter of weeks than they had previously been able to accomplish in months or even years. Government has generated new data sets and data collection processes and joined up existing data in new ways to meet specific and urgent needs. In our report on *Initial learning from the government's response to the COVID-19 pandemic*, we cited the following examples:[9]

- Ministry of Housing, Communities & Local Government (MHCLG) – introduced a survey of monthly cost pressures and income losses in local authorities;
- MHCLG – collected data on the potential scale of the population in England which was either sleeping rough or was at risk of doing so; and
- NHS England & NHS Improvement – commissioned a tool for care homes which was adapted and expanded to collect a wide range of additional data beyond numbers of vacant beds.

These are examples of non-personal data being aggregated for beneficial outcomes. However, personal data is often more challenging. The compilation of the clinically extremely vulnerable list during lockdown is an example of the issues faced when trying to work with personal data within the limits of existing systems. This exercise needed to access, extract and combine data from multiple sources and highlighted the difficulties of doing this in a systematic way (**Figure 3**).

The events of 2020 showed that people can come together quickly at a time of national emergency and achieve much in a short space of time. However, such conditions create the risk that activities are essentially one-off exercises delivering tactical solutions that are not repeatable or sustainable. They can also make it harder to achieve the overall strategic aims in the long run.

Data-quality problems will persist unless the more fundamental issues are addressed. These include legacy data estates, funding, a perception of conflict between immediate delivery and quality, siloed working and the more general points made in the introduction about a culture of tolerating and working around poor-quality data.

### Figure 3

Case example: Identifying the clinically extremely vulnerable during the COVID-19 pandemic, 2021

**Significant effort remains in combining data sets at scale across government**

**Main department:** NHS Digital, Government Digital Service and the Ministry of Housing, Communities & Local Government (now Department for Levelling Up, Housing & Communities)

**Objective:** The shielding programme was a swift government-wide response to identify and protect clinically extremely vulnerable people against COVID-19. To achieve this the government needed to bring together data from existing, separate data sources to urgently identify the people who were clinically extremely vulnerable.

**What happened:** At the start of the pandemic, there was no mechanism to allow a fast 'sweep' across all patients to identify, in real-time, those who fell within a defined clinical category. NHS Digital developed the list in several iterations, as more data became available. The first iteration, based on hospital, maternity and prescribed medicines data, was ready on 20 March 2020. The second iteration, using GP patient data, was released on 12 April owing to the time needed to extract this data as NHS Digital did not have ready access to this data set. It took NHS Digital three weeks to carry out the technical task of accessing and extracting GP patient data.

**Outcome:** The government identified lessons from the first iteration of shielding and sought to apply them to the second lockdown towards the end of 2020. However, we reported that during the second lockdown systems remained incapable of speaking to each other, although the government set up a new national shielding service system designed to improve its ability to view and analyse data on clinically extremely vulnerable people and their needs.

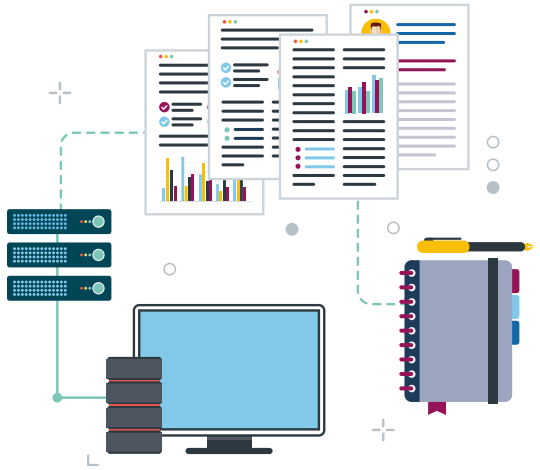
**What lessons departments can learn:** There are significant constraints that need sustained effort to overcome, which apply to all areas of government trying to use and share data beyond its original purpose. The government needs to address the issue in a managed and incremental way, rather than resorting to one-off exercises, which departments must repeat manually.

Source: Comptroller and Auditor General, *Protecting and supporting the clinically extremely vulnerable during lockdown*, Session 2019–2021, HC 1131, National Audit Office, February 2021

9 Comptroller and Auditor General, *Initial learning from the government's response to the COVID-19 pandemic*, Session 2021-22, HC 66, National Audit Office, May 2021.

# Data standards

## Why are standards so hard to implement?



Where standards are not adequately defined different stakeholders may implement them in different ways, with the result that their systems do not achieve the interoperability intended because the data may not be interpreted in a consistent manner. For example, in our report *Rolling out smart meters*, we found that despite government setting technical standards for industry to follow, the use of multiple suppliers and manufacturers caused problems with design and interoperability and increased the complexity and delay of the rollout.[10]

Government has long recognised the role of standards in making it easier to link and share information across organisational boundaries. But the structure of government is heavily siloed and departments have a high degree of autonomy. With data also siloed, joining it up across different systems and organisations is a difficult challenge.

This is another area where despite the intent, government has struggled to make substantial progress. In 2019, nearly 20 years after government first set out the need for a common standard format for citizen name and address, we published *Challenges in using data across government*. Among 10 departments and agencies we found more than 20 different ways of identifying individuals and businesses, with no standard format for recording data such as name, address, and date of birth. The problem is replicated for businesses and in local areas where information is recorded differently across boundaries. This makes it difficult for government to connect its data across different sectors to analyse and understand economic challenges or systemic problems.

10 Comptroller and Auditor General, *Rolling out smart meters*, Session 2017–2019, HC 1680, National Audit Office, November 2018.

## Legacy systems make it difficult to introduce standards into an existing environment

Failure to appreciate and understand the legacy environment and the practical issues it creates for data-sharing is one of the biggest barriers to using and sharing data. 'Legacy' refers to the people, processes, systems and data which no longer meet business needs and are constrained by old technology. It is not cost-effective to replace all systems whenever a new need or a better technology is identified. Indeed, replacing embedded legacy systems is in itself a risk and in some cases the risk may outweigh the benefits. 'Lifting and shifting' legacy systems to cloud hosting can address hardware obsolescence but does not in itself address systems and data issues.

Legacy systems can be a barrier to digital transformation, and it may not be feasible to retrofit data standards to legacy systems. Established systems may continue in use for many years and can be expensive to replace. For example, in our report *Digital transformation in the NHS* we said that the typical replacement cycle for an electronic patient records system can be up to 15 years and the cost of replacement can be upwards of tens of millions of pounds for each organisation.[11]

'Technical debt' is the estimated cost of future development to make a service or product function optimally. It can arise for a variety of reasons including taking a tactical approach for expediency rather than a better approach for the longer term but which would take more time upfront. The Cabinet Office has initiated a remediation programme across government to tackle legacy and technical debt. It has cautioned that departments should ensure that technical choices such as building integrations between systems to improve interoperability and data insights do not come at the expense of creating, extending or embedding reliance on legacy systems.

11 Comptroller and Auditor General, *Digital transformation in the NHS*, Session 2019–2021, HC 317, National Audit Office, May 2020.

# Resourcing

## Why can't we put more resources into tackling the problem?

Data projects are often set aside when funding is under pressure. Schemes to provide basic data improvements do not always receive investment. For example, in our *Challenges in using data across government* report, we reported that the Department for Business, Energy & Industrial Strategy produced a business-wide data strategy but did not receive funding to help make it possible. It can be difficult to make the case for funding stand-alone data projects, such as to build a data model, define data standards, or improve the quality of data.

Departments rarely measure the inefficiencies in terms of time and cost that arise from the need to match, clean, combine and improve poor-quality and disorganised data. The Government Data Quality Framework states that the costs of fixing data-quality problems are often seen to be too high, but the costs of not fixing quality problems are underestimated and these should also be understood.

Sometimes, the costs of cleaning up or matching data fall to one organisation while the benefits are felt by another. However, civil service funding and performance-monitoring arrangements hold government departments to account individually. This means they do not readily support cross-departmental working where one or more departments need to contribute resources to support another department to achieve its objectives. This runs counter to government's aims for greater sharing of data across departmental boundaries.

In our report *Challenges in using data across government* we pointed to successful cross-government working by HMRC and DWP on RTI. This was a complex project in which data architects from HMRC worked with DWP for several years to make the data systems compatible. It shows how policy imperatives, senior-level commitment, collaborative working and continued commitment can achieve a positive outcome despite the difficulties and complexities.

Our report highlighted the following factors which would help with aligning the funding requirements:

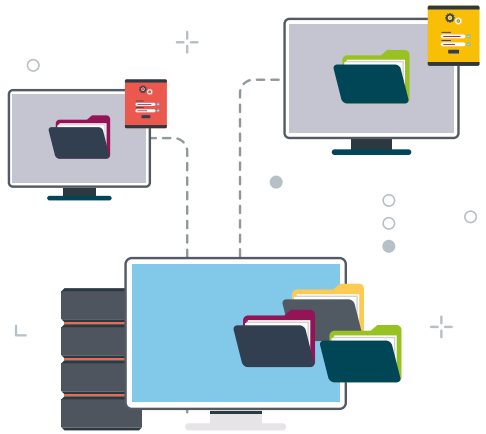
- An understanding of how departments or public bodies use the data, and which will benefit most from change – this could lead on to identifying which departments have the greatest incentive to lead the work, where the greatest burdens will fall and which bodies will need the most support.
- A mechanism for funding cross-government working which provides a good return on investment for government as a whole in the long term, even though some departments that are crucial to success may have less to gain.

Our report *Financial modelling in government* gives further examples of how departments assure the quality of data received from other parts of government in the context of creating business-critical models for decision-making.[12]

12 Comptroller and Auditor General, *Financial modelling in government*, Session 2021–22, HC 1015, National Audit Office, January 2022.

# Access to raw data and APIs

## Why can't departments just give each other access to the raw data or access it through application programming interfaces (APIs)?



We have seen examples of API use in government using key fields to check a data source. For example, DWP provides an 'NHS charge exemption' API which can be used as one of the checks to determine whether a person is eligible for free prescriptions.

But the answers to the question need to take account of all the issues set out above, particularly around data quality. In our report *Challenges in using data across government*, we said that:

- departments suggested that between 60% and 80% of time is spent cleaning and merging data. In some areas this can equate to several hundred analysts' time;
- while it may be technically feasible to retrofit APIs to existing systems, it can be a difficult and expensive process to do so; and
- APIs may perform poorly under load, especially where there is a high throughput, or where data is not indexed in the source system.

Providing access to raw data without further information about that data means potential users may not understand important context relating to quality, such as when the data was collected and last updated, and any limitations in collection of data. Without such understanding, people may misuse the data leading to poor outcomes and reputational damage. Likewise with an API call, the organisation using the data needs a means of establishing confidence that the data returned is sufficiently reliable for its intended purpose.

# Creating cross-government data sets for multiple users

## Why can't government create single data sets from each other's data?

Where data is brought together from disparate sources to create a new data set, further questions arise:

- ownership – who owns it on an ongoing basis?
- maintenance – who keeps it up to date on an ongoing basis?
- funding – who pays for it on an ongoing basis?
- privacy – who is responsible for privacy and consent?
- risk – how does the aggregation of data impact risk appetite?

Where these questions are not addressed, bringing data together can be a one-off exercise rather than being sustainable or repeatable. CDDO acknowledges that the question of risk in particular is a real concern and is a reason why data-sharing initiatives should consider alternative approaches such as data trusts and data virtualisation.

The *National Data Strategy* included an action for the Office for National Statistics (ONS) and Cabinet Office to create an Integrated Data Service (IDS), as a key component of improving the analytical evidence base for policy development. The IDS will include the creation of a series of integrated data assets, which will combine data from across the public sector relevant to key priority areas to enable complex analyses and insights to be drawn in order to inform policy development and service provision. ONS is leading a programme to develop and run the IDS, drawing heavily on its experience running the Secure Research Service, which makes de-identified data available to researchers, and in creating and analysing complex integrated data assets.

# Data analytics

## Why can't data analytics solve the problem?

In our report *Challenges in using data across government* we said that departments suggested that between 60% and 80% of analysts' time is spent cleaning and merging data, even for internal uses. This situation is not unique to the public sector. The question therefore comes back to a proper assessment of data quality. Where this is assessed as good, data may be suitable for sharing with no further action required. Or it may be possible to package the data for use by others after a modest level of cleansing. But there will be situations where the condition of the data will make analytics difficult to apply, for example where there is a high prevalence of freeform notes and other conflicting formats.

As we said in our opening comments, those organisations which are best able to take advantage of advanced data technologies are those which were created to do so from the outset or have successfully overcome the challenge of legacy.

# The way forward

## Reviewing why initiatives have failed in the past

Government's strategies often aim for 'better use of data' as if it is low-hanging fruit with the benefits there for the taking. But they fail to learn from the multiple times government has struggled to translate this vision into successful implementation. To move forward government needs to acknowledge and address the difficulties we highlight here and understand why previous initiatives have not succeeded. Despite this observation, government now has another new strategy for data. This 2020 National Strategy sets out an ambition to drive change across several key areas for data use in government (**Figure 4**).

### Figure 4

Transforming government's use of data to drive efficiency and improve public services

| Area for change | Aim of the National Data Strategy |
|---|---|
| Quality, availability and access | Better-quality data that is collected and held consistently and with clarity and efficiently shared between organisations |
| Standards and assurances | Adoption of standards for data, leading to greater consistency, integrity and interoperability, with data used widely and effectively across government |
| Capability, leadership and culture | Capability in data and data science across central and local government, so leaders understand its role, expertise is widely available, staff at all levels have the skills they need, and a cross-government 'data-sharing by default' approach tackles the culture of risk aversion around data use and sharing |
| Accountability and productivity | Opening government up to greater scrutiny and increasing accountability, ensuring that this drives improvements in productivity, policy, and services for people, while also ensuring data security; and using procurement to drive innovation and better outcomes |
| Ethics and public trust | Transformation will only be possible and sustainable if developed in a robust ethical framework of transparency, safeguards and assurance which builds and maintains public trust in the government's use of data |

Source: National Data Strategy 2020, Mission 3: Transforming government's use of data to drive efficiency and improve public services (available at: www.gov.uk/government/publications/uk-national-data-strategy/), link accessed 12 July 2022

# Embedding data standards

The need to establish common standards and infrastructure to enable interoperability across government, including the definition and adoption of standards for common data such as citizen name and address, was first set out in 2000 and shows the depth of early thinking.[13] Subsequent strategies have had similar aims.[14] However, implementation has yet to be achieved.

## Where to start with data standards

A look at the costs and time spent on working around discrepancies in systems should provide insight on how and where to focus resources. Examples could be where names vary across systems, where National Insurance numbers may not match or where data is recorded differently in differing local systems.

In our *Challenges in using data across government* report we set out the following practical steps for government to take for standardising data:

- Identify key government data sets.
- Gain cross-government consensus on key data fields for government as a whole.
- Set data standards for all key fields. These should be consistent across all datasets and should be used for all new systems.

The Committee of Public Accounts has urged Cabinet Office to identify and prioritise the top 10 data standards of benefit to government.

## Pitfalls to avoid with data standards

Once standards are clearly defined, they need to be adopted consistently. In our report *Digital transformation in the NHS*, we said that interoperability had been made more difficult by previous attempts to implement standards. These resulted in the use of multiple standards, including alternative standards for the same function and several generations of the same standard being in use at the same time. We said that if the NHS does not develop and implement a carefully considered plan, then it risks not only failing to take the right steps towards interoperability in the short-to-medium term, but also making it harder to achieve interoperability in the future.

We welcome the creation of the following cross-government initiatives:

- CDO Council – set up in November 2021 for chief data officers and data leaders to steer strategic execution and promote collaboration.
- Data Standards Authority – created in April 2020 to establish standards to make it easier and more effective to share and use data.
- Data Architecture Design Authority – a new body to review, approve and monitor adoption of data architecture principles and frameworks.

13 Cabinet Office, *e-government – A strategic framework for public services in the information age*, April 2000, (available at: https://webarchive.nationalarchives.gov.uk/ukgwa/20031020010535/http://www.e-envoy.gov.uk:80/EStrategy/StrategicFramework/fs/en, link accessed 12 July 2022).

14 For example, Cabinet Office, *Government ICT Strategy – Smarter, cheaper, greener*, January 2010, (available at: https://webarchive.nationalarchives.gov.uk/ukgwa/20100304104621/http://www.cabinetoffice.gov.uk/cio/ict.aspx, link accessed 12 July 2022).

# Taking a structured approach

There are a variety of frameworks and maturity models available for organisations wishing to improve their quality and use of data (**Figure 5**). This guide does not seek to replicate them but aims to give an indication of what is available and how it can be used.

If government is to improve its data, it will need to actively manage that data. However, not all departments have chief data officers or data managers and government's current Digital, Data and Technology skills framework does not define the role of the data manager.[15] This is in contrast to the globally accepted Skills Framework for the Information Age, which is used as a framework in the wider profession and does include data management as a skill.[16]

## Figure 5

Data frameworks and maturity models

| Framework or model | Purpose and coverage |
|---|---|
| Government Data Quality Framework | Focuses mainly on assessing and improving the quality of data input within central government. It is in two parts:<br>• Part One: provides a structure for organisations to frame their thinking around data-quality principles, the data lifecycle and data-quality dimensions against which data quality can be assessed.<br>• Part Two: provides guidance on techniques to assess, communicate and improve data quality, including the use of data-quality maturity models and action plans. |
| Government Data Maturity Model | This is under development with a pilot being rolled out from December 2021. It has not been publicised more widely at the time of writing this guide. The Framework notes that there are various maturity models with slightly differing themes, and gives the following as an example:<br>• Leadership and culture.<br>• Skills.<br>• Tools and architecture.<br>• Data governance.<br>• Quality and standards. |
| Environment Agency Data Integrity Maturity Model | Mentioned in the Government Data Quality Framework as a case study example of an organisation which has applied a maturity model to assess and improve its data over time. It contains the following headings:<br>• Governance and accountability<br>  ◦ Governance<br>  ◦ Ownership<br>  ◦ Security<br>• Line of sight<br>  ◦ Sharing<br>  ◦ Interdependencies<br>• Data standards and quality monitoring<br>  ◦ Data standards<br>  ◦ Data quality and confidence |

Source: Government Data Quality Framework, (available at: www.gov.uk/government/publications/the-government-data-quality-framework); Environment Agency Data Integrity Maturity Model, (available at: https://defradigital.blog.gov.uk/wp-content/uploads/sites/136/2016/04/16-08-05-Model-FINAL.pdf), links accessed 12 July 2022

15 See: https://sfia-online.org/en/tools-and-resources/standard-industry-skills-profiles/uk-government-ddat-roles/sfia-skills-profiles-for-uk-ddat-roles, (link accessed 12 July 2022).

16 See: https://sfia-online.org/en/sfia-8/skills/data-management, (link accessed 12 July 2022).

The CDDO and ONS Data Quality Hub have developed a Data Management Maturity Model tailored to government based on a model developed by Data Orchard. CDDO is piloting it in six departments and expects to roll it out more widely during 2022-23. The expectation is that this will provide a universal assessment of government's data management maturity for future benchmarking work.

CDDO has also published a Data Governance Sharing Framework, which sets out a set of principles and underpinning actions that can help organisations address the non-technical barriers to data-sharing.[17]

In our report *Challenges in using data across government*, we published an analysis of a sample of departmental data strategies and documentation in the following areas:

- Strategy
- Business alignment
- Governance

The criteria we used are set out opposite.

### Strategy

What should be in place (at a minimum):

- The organisation has an enterprise-wide data strategy.
- The data strategy is consistent with the organisation's overall objectives.
- The data strategy has active business ownership and commitment.
- There is a road map for data improvements.
- The ongoing and development costs of data services are understood.
- Data changes and migration activities are planned and costed into business cases.
- The burden on front-line staff and service users to capture quality data is recognised and managed.
- The cost of poor-quality data has been quantified.

Questions to ask:

- **Strategy:**
  - Are there clearly articulated business goals, with aligned underpinning strategies for technology and data, supported by active business commitment and ownership?
- **Funding and burden:**
  - Is there a data strategy with a clear road map of data improvements, with prioritised and funded projects?
  - Is the burden of collecting quality data recognised and managed/minimised, and ideally quantified?
  - Are data needs including cleansing and migration fully planned and costed in business cases for change?

17 Central Digital and Data Office, *Data Sharing Governance Framework*, May 2022. Available at: www.gov.uk/government/publications/data-sharing-governance-framework, (link accessed 12 July 2022).

## Business alignment

What should be in place (at a minimum):

- Data is treated as a strategic asset.
- Information flows easily to where it is needed to support business activities.
- There is a shared data model.
- The data model is recognised and supported by all business areas.

Questions to ask:

- **Information alignment:**
  - Is information recognised across the organisation as a strategic asset?
  - Is there a clear alignment between the strategy and the information needs?
  - Do business areas have appropriate input into the development of the data strategy, data architecture and enterprise data model?
- **Data to support the business:**
  - Is timely information provided to support strategic/operational decision-making?
- **Data model and analysis:**
  - Is there a shared data model across the organisation which is recognised and supported by all business areas?
  - Is data accessible for reporting and further analysis?

## Governance

What should be in place (at a minimum):

- The organisation has a data architecture.
- The data architecture is integrated into the wider enterprise architecture.
- There are clear data governance rules, accountability and ownership of data sets.
- Data quality is monitored and assessed against suitable metrics.
- The organisation understands and manages its legal obligations relating to data protection and sharing.
- Governance arrangements provide strong and effective oversight.

Questions to ask:

- **Data structures, rules and definitions:**
  - Is there a data architecture, consistent with a wider enterprise architecture (high-level view of how the organisation fits together)?
  - Are there clear rules around data governance, accountability and ownership?
- **Degree of trust/data quality:**
  - Do users trust the information they receive?
  - Is quality monitored and assessed with a coordinated approach to continuous improvement?
- **Legislation**:
  - Does the organisation manage compliance with legal obligations around the collection, storage and use of data (Digital Economy Act 2017, Data Protection Act 2018)?
- **Governance framework:**
  - Are there arrangements to provide strong and effective oversight?
  - Do key stakeholders understand the value of their role and are they empowered to perform it?

## Addressing legacy issues

To improve its legacy systems and remediate them effectively government must first identify what it has under management across the sector.

In its report on the *Challenges in using data across government*, the Committee of Public Accounts recommended that the Cabinet Office and DCMS should:[18]

- identify the main ageing IT systems that, if fixed, would allow government to use data better; and
- ensure that whenever departments replace or modify these systems this is done with full consideration of how the systems will support better use of data in government.

## Enabling data-sharing

### Risk assessments for data-sharing

The Open Data Institute (ODI) has recently published *Assessing risks when sharing data: a guide*.[19] This aims to help organisations, not only those in government, identify potential data-sharing issues and contains guidance on undertaking a data-sharing risk assessment. The areas and high-level questions are in **Figure 6**. It may help government organisations to recognise that sharing data may create risks because of limitations in their processes or analysis. The damage caused if these risks are not recognised is exemplified by our report on *Handling of the Windrush situation* where the department concerned shared data without fully assessing its quality with the potential for citizens being wrongly detained, removed or denied access to public services.[20]

The ODI guide recommends that data should be shared with well-structured and high-quality documentation and metadata to help new users understand important context, such as: its quality, when it was collected and last updated, and any limitations. This will help them understand whether and how far they can make use of the data and can help mitigate any risks of misuse of data and potential reputational damage.

Where the results of a data-sharing risk assessment indicate that data is not suitable for sharing, there is value in explicitly articulating the reasons. This can help in understanding the barriers to re-use of data and what can be done to address them.

18 HC Committee of Public Accounts, *Challenges in using data across government*, One Hundred and Eighteenth Report of Session 2017–2019, HC 2492, September 2019.

19 Open Data Institute, *Assessing risk when sharing data: a guide*, February 2022. Available at: https://theodi.org/article/assessing-risk-when-sharing-data-a-guide/, (link accessed 12 July 2022).

20 Comptroller and Auditor General, *Handling of the Windrush situation*, Session 2017–2019, HC 1622, National Audit Office, December 2018.

## Figure 6

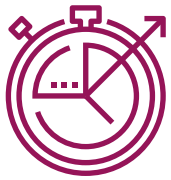
Risks and questions to consider when data-sharing

| Category | Perceived or actual risks | Key questions to consider |
|---|---|---|
| Legal and regulatory | Breaching data protection law, intellectual property rights, other regulatory requirements, or legal contracts | 1 Does the data contain any personal data?<br>2 Does the data contain third-party data?<br>3 Do you have the legal permissions to share the data?<br>4 Are there any other relevant legal considerations? |
| Ethical | Enabling unethical data collection or use, or directly impacting people and communities | 5 Are there any relevant cultural considerations?<br>6 Is sharing the data likely to impact people or communities?<br>7 Will sharing the data impact the natural environment?<br>8 Does the data contain anything that could impact national security?<br>9 Does the data contain anything that could impact the security of the organisation or its staff? |
| Reputation | Suffering reputational damage from sharing or using data that breaches trust, or that reveals limitations in processes or analyses | 10 Will anyone be surprised by you holding, sharing or using this data?<br>11 Is a data-quality caveat needed?<br>12 Are there any free-text or comment fields in the data set? |
| Commercial | Losing competitive advantage in the market | 13 Does the data contain anything commercially sensitive? |

Source: Open Data Institute, *Assessing risk when sharing data: a guide*, February 2022

# Conclusion

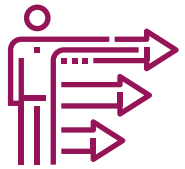
Our key messages are:



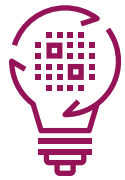
Government data is a leading cause of inefficiencies

Organisations need to acknowledge the importance of fixing underlying data issues

Data needs to be managed and this cannot be achieved without focused effort, funding and prioritisation



Initiatives tend to peter out when the going gets tough, but it is important to continue



Fully addressing the issues highlighted in this guide could bring substantial benefits by enabling reform and transformation

# Appendix One

## References

*National Data Strategy 2020*
www.gov.uk/government/publications/uk-national-data-strategy/

*Government Data Quality Framework*, December 2020
www.gov.uk/government/publications/the-government-data-quality-framework/

Open Data Institute, *Assessing risk when sharing data: a guide*, February 2022
https://theodi.org/article/assessing-risk-when-sharing-data-a-guide/

*Data Sharing Governance Framework*, May 2022
www.gov.uk/government/publications/data-sharing-governance-framework

Central Digital and Data Office, *Transforming for a digital future: 2022 to 2025 roadmap for digital and data*, June 2022
www.gov.uk/government/publications/roadmap-for-digital-and-data-2022-to-2025

# Appendix Two

## Progress implementing the NAO's recommendations

The Department for Digital, Culture, Media & Sport, which is responsible for drafting the National Data Strategy and for data policy at the national level, across society and the economy, and the Cabinet Office, which is responsible for the government's use of data, should:

| | Recommendation text | Implementation status | Actual/Expected implementation date |
|---|---|---|---|
| 1 | Use the data strategy to identify and address the barriers to better use of data. It should include a clearly articulated plan of work to overcome these barriers. This should provide an assessment of fundamental data issues, including safeguarding data and public trust, and plans for improving the communication of government's approach, and potential benefits of using data more effectively. | Implemented | March 2021 |
| 2 | Set up clear cross-government accountability, governance and funding for data to support delivery of the data strategy. Joint working and cross-government groups need to have clearly assigned responsibilities that are aligned with the levers available including funding, controls and operational resources. These arrangements should be clearly communicated across government to alleviate confusion of where responsibilities lie. | Implemented | March 2020 |
| 3 | Develop cross-government rules, standards and common ways to collect, store, record and manage data. Where multiple standards are used, government should develop a consistent approach to balancing competing demands between standardisation and local requirements, including implications for future decision-making and costs. This should include a regular review of departments to ensure that they are applying these standards and principles to their data collection. | Work in progress | End of Spending Review 2021 |
| 4 | Identify data sets that are critical to government functions, look at how to share them easily and examine how they can be enhanced by process improvement and automation. This should include an analysis of the processes, systems and data flows so their use is fully understood. | Implemented | June 2022 |
| **Other departments which are different levels of maturity should:** | | | |
| 5 | Put in place governance for data, including improving executive team understanding of the issues associated with their underlying data and the benefits of improving their data. | Work in progress | End of Spending Review 2021 |
| 6 | Set out data requirements in business cases. This should include an assessment of the current state of the data, implications for confidence in spending decisions, and the improvements or new data that are needed to support implementation of the project. These assessments should have an explicit consideration of the ethics and safe use of the data under discussion. | Work in progress | End of Spending Review 2021 |
| 7 | Implement guidance for front-line staff for handling data. This needs to recognise the effort and resource required to fully and consistently adopt the policy and principles created by government into the working practices of the department, including standardisation, data ethics and quality. | Work in progress | End of Spending Review 2021 |

**Note**

1 Spending Review 2021 sets departmental budgets up to 2024-25.

Source: National Audit Office recommendation tracker, (available at: www.nao.org.uk/nao-recommendations-tracker/), updated periodically. This reflects the situation at the time of publication of this guide

# Appendix Three

## Previous NAO reports

In 2021 we reported on the *Challenges in implementing digital change* and reiterated our finding that government has yet to address the underlying barriers and constraints that make the ambition to join up data such a difficult undertaking to implement in practice. There is still a poor appreciation of the state of the data in legacy systems and its impact on the transformation of operational services. Data issues include age, quality and consistency across different systems.

## What we have found through our work

We have published a number of reports in recent years which highlight that there is still much work to do on improving the use of data across government.

Findings from *Digital transformation in government* (2017):

- Previous attempts to map the data environment failed because of the fragmented landscape and burden of detail.
- Little central strategic overview of the data needs of departments.
- Many important and difficult aspects of data use are still to be addressed.

Findings from *Challenges in using data across government* (2019):

- Data is not always seen as a priority, with quality and sharing of data clear examples of neglected and poorly planned activities.
- Quality of data not well understood, with new initiatives exposing the poor quality of the underlying data.
- Culture of tolerating and working around poor-quality data.

Findings from *Digital transformation in the NHS* (2020):

- Previous attempts to implement standards have resulted in the use of multiple standards or different versions of the same standard.
- Risks that current initiatives could make interoperability harder to achieve in the future.
- Difficulty and expense of modifying existing legacy systems to accommodate new standards.

Findings from *The challenges in implementing digital change* (2021):

- Despite a high-level acknowledgement that data is a key asset, government still has a poor appreciation of the state of the data in legacy systems and its impact on the transformation of operational services.
- Government transformation programmes and business cases often fail to explicitly address data at the start, and instead it becomes an area of concern and delay further into the project.
- Government has an ambition to join up data but has not yet addressed the underlying barriers and constraints that make this such a difficult undertaking.

We focused our review on the use of data to support delivery of public services, but many of our findings are equally relevant to data to support decision-making and improve performance, for example in finance, HR and shared services.