Office for Statistics Regulation

Systemic Review Programme

Joining Up Data for Better Statistics

September 2018

Office for Statistics Regulation

We provide independent regulation of all official statistics produced in the UK. Statistics are an essential public asset. We aim to enhance public confidence in the trustworthiness, quality and value of statistics produced by government.

We do this by setting the standards they must meet in the Code of Practice for Statistics. We ensure that producers of government statistics uphold these standards by conducting assessments against the Code. Those which meet the standards are given National Statistics status, indicating that they meet the highest standards of trustworthiness, quality and value. We also report publicly on system-wide issues and on the way statistics are being used, celebrating when the standards are upheld and challenging publicly when they are not.

Acknowledgements

We would like to acknowledge the significant input of all the individuals and organisations that contributed to this Review as we gathered information and tested the ideas presented here. Their details are listed in Annex 2.

Contents

Executive summary

Why we did this review

ES.1. In 2017, the Office for Statistics Regulation launched an investigation of the UK statistics system's ability to provide greater insight to users via data linkage. We are the independent regulators of the UK's official statistics system and our interest in this area is underpinned by the Code of Practice for Statistics' three pillars of trustworthiness, quality and value:

- Statistics add **value** when they answer society's questions. Many questions cannot be answered without sharing and linking data. As a result, a greater willingness and ability to share and link data is an essential prerequisite for improved official statistics.
- Without a focus on the **quality** of the data – their source, how they have been collected and processed, any biases and incompleteness in the data – the results could be misleading.
- Custodians of public data must demonstrate their **trustworthiness** by safeguarding data robustly during and after the sharing and linkage process, and by being open to public scrutiny. Organisational trustworthiness is at the core of OSR's work and is a key component of the first pillar in the Code of Practice

What we found

ES.2. Data linkage should be a vital component of the official statistics landscape. Value is being squandered because this is not currently the case. New powers have been given to the statistics system via the Digital Economy Act (DEA) 2017 that have the potential to unlock new sources of data and transform the official statistics landscape. Plans for new investment in external researchers' use of administrative data are also due to be announced soon. Users' aspirations within and beyond government are high and their demand is clear: make this work.

ES.3. There are some powerful examples of data linkage being used in government to provide insights and drive policy change. This report highlights many such projects, and we know there are more. For example, data linkage means that prospective undergraduate students can now compare the long-term career and earning prospects for different courses across the UK. Organisations delivering interventions to reduce reoffending and increase employment rates for prisoners in England after they are released can assess their effectiveness. Analyses of the impact of decisions made in the family courts in England and Wales on children's long-term educational outcomes will be used to shape future policy. Scotland is joining-up data from multiple sources to inform its strategy to reduce deaths from cardiac arrest, including guiding decisions about where to locate community defibrillators.

ES.4. However, these success stories remain the exception, and the concern is that the efforts and time required to create these linked data resources can discourage others seeking to do similar work. Government analysts face frustrating conversations with colleagues to explain why data cannot be joined-up to evaluate policies in useful timeframes (or at all). Many parts of government lack effective mechanisms for identifying important cross-cutting questions that could be answered through data linkage. Risk averse data controllers and the need for new legal gateways result in data shares taking years to progress. Users face an array of different approval processes to navigate. Data are often poorly documented so their full potential lies untapped and quality assessment processes are hindered. Not enough analysts have the expertise to carry out linkage and analyse the resulting data, and the people within the system with the capability and desire to share expertise do not have spare capacity to do so.

ES.5. The DEA is a welcome development, but it cannot alone address all the barriers to effective data use that we heard about, including those experienced by the Administrative Data Research Network over the past five years, and health and social care data are excluded from its research provisions. We have identified six outcomes designed to ensure that the UK has a safe and effective data linkage system to support statistics production and statistical research. At the core, underpinning all other outcomes, is a system that understands the importance of demonstrating its trustworthiness as a

custodian of public data. This demonstration of trustworthiness should serve as a platform to support greater use of data sharing and linking to deliver insights about important issues. The six outcomes, and our recommendations for action to achieve them, are set out below.

ES.6. OSR is one of many organisations with an interest, and part to play, in delivering the changes required to ensure that data sharing and linkage are used to maximum effect for statistics production. The key to success will be harnessing and coordinating all available opportunities for leadership on data sharing and linkage issues.

Achieving a safe and effective data linkage system

1. Actively seek input from the public in major decisions about statistics production and statistical research using data linkage.
2. Identify clear, consistent and meaningful language to use when engaging with the public about data safeguarding, linkage and use.
3. Be advocates for safe data use to provide insights that serve the public interest.
4. Produce "keeping data safe" statements using the Five Safes Framework.
5. The Department for Digital, Culture, Media and Sport should develop an overarching framework for trustworthy government data use.

6. Maximise opportunities to identify the questions that society wants answered by exploiting existing networks of senior leaders e.g. Heads of Profession, Chief Statisticians, Directors of Analysis and Chief Scientists.
7. Ensure that policy makers and external experts are actively involved in processes to identify questions.
8. Ensure that departmental Areas of Research Interest Statements explore cross-cutting interests across departments, and are produced with input from all analytical professions.
9. Enable more exploratory analysis to take place before research questions are finalised, including through the use of synthetic data.

10. Agree common information governance frameworks to harmonise practice across government departments.
11. Consider placing the responsibility for signing off data sharing agreements with more senior staff.
12. Explore the contribution that risk assessment tools could make to decision-making about data shares.
13. Publish materials related to data shares, including mandatory and voluntary Data Protection Impact Assessments, to support transparency.

14. Design data access application processes and support materials with user input and seek ongoing feedback on systems when they are live.
15. Signpost users to other sources of data of potential interest to them.
16. Work with health data users and NHS Digital to scope the terms of an independent review of NHS Digital's health data sharing and access processes.

17. Identify data sources most in need of additional documentation and metadata and work with data owners to address gaps.
18. Create a central registry of administrative data sources.
19. Ensure that each step in the data linkage process is documented using a common framework.
20. Ensure that statistical experts are consulted and participate in the design process for new data systems.
21. Identify mechanisms for data users to feedback information about data quality to data collectors to help improve data at source.

22. Recognise resource needs – including the imbalance of demands placed on data holding departments – and either address them with additional inputs, or be clear about what constraints are faced and their implications.
23. Government departments to work together to identify resource-efficient solutions to infrastructure requirements (e.g. data storage space, software).
24. Identify creative solutions to cut the cost of data extracts charged by external contractors; address this issue in any future contracts.
25. Develop a new data linkage skills strategy to support the expansion of opportunities for training and development in this area.
26. Identify effective mechanisms to bring in external expertise on data linkage methods and analysis from academics and other experts.
27. Ensure the professional development needs of staff who support the data access process are met, including opportunities to network and share practice with people in other organisations in similar roles.
28. Continue to innovate and share practice around the delivery of safe data settings, especially virtual solutions.
29. Develop a network of accredited safe settings with common operating standards to act as a single entry point for data users.

Chapter 1: Introduction

Why we did this review

1.1. In 2017, we launched an investigation of the UK statistics system's ability to provide greater insight to users via data linkage. As the independent regulator of the UK statistics system, OSR has a major interest in the sharing and linking of data for the reasons set out below. These draw on the trustworthiness, quality and value framework in the new Code of Practice for Statistics:

- Statistics add **value** when they answer society's questions. Many questions cannot be answered without sharing and linking data. As a result, a greater willingness and ability to share and link data is an essential prerequisite for improved official statistics. OSR's focus is on improving the public value of statistics, and we can use our voice to emphasise the importance of linking to the production of official statistics.
- Without a focus on data **quality** the value of linked data could be overestimated and the results disappoint (or worse, be misleading). OSR has developed a framework for data quality, and rolled out the Quality Assurance of Administrative Data approach. We can play an important role in highlighting to statisticians and other analysts that they should not neglect quality in their enthusiasm to link datasets.
- Custodians of public data must demonstrate their **trustworthiness** by safeguarding data robustly during and after the sharing and linkage process, and by being open to public scrutiny. Organisational trustworthiness is at the core of OSR's work and is a key component of the first pillar in the Code of Practice.

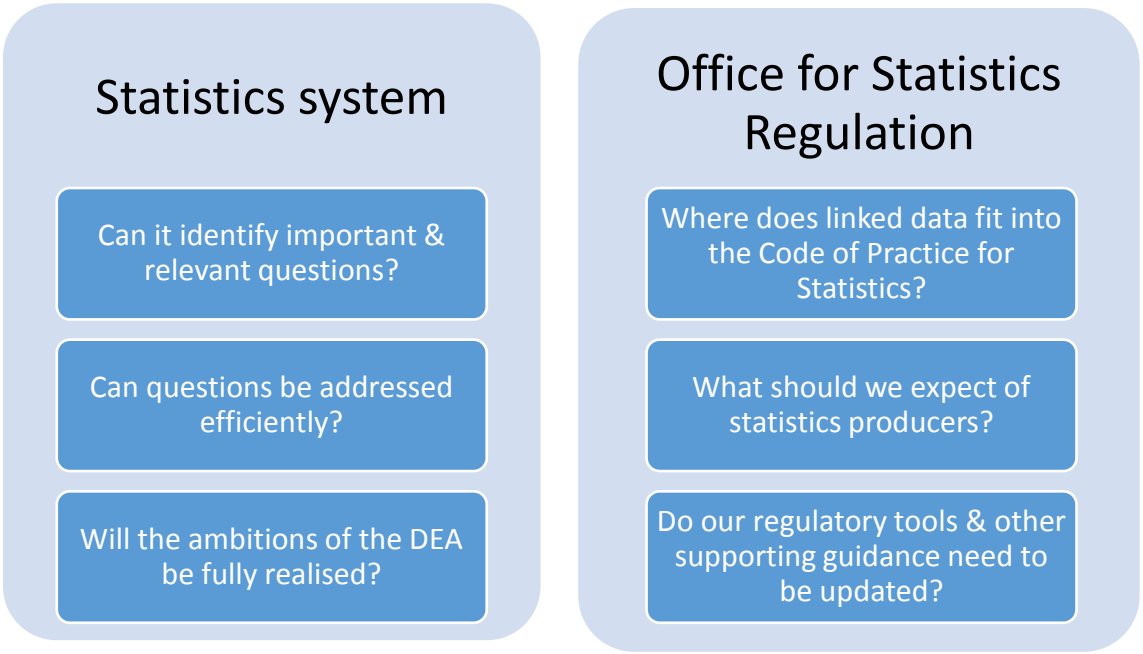
1.2. Why now? We wanted to understand how the statistics system was preparing to respond to the Digital Economy Act (DEA) 2017, and wanted a firm basis for our ongoing monitoring of the implementation of the Act from 2018 onwards. The DEA amends the Statistics and Registration Service Act 2007 to support the UK Statistics Authority's statistical production functions (which are carried out by the Office for National Statistics). It creates a legal basis for the Authority to access data held by Crown bodies, other public authorities, private organisations and charities.

1.3. Greater access to these sources of data can support the increased use of data linkage, creating opportunities for the statistics system to provide new insights about many areas of life. In some situations, data sharing and linkage also has the potential to reduce burden by replacing direct data collections from people and organisations with analyses of administrative data collected in the course of government providing services.

1.4. External users with an interest in accessing government data for research studies will be supported by the DEA's research provisions. These enable public bodies to make data available to researchers for research that is in the public interest, providing the proposed studies and researchers are accredited and the data are safeguarded to minimise the risk of an individual's identity being disclosed. The UK Statistics Authority is the accrediting body for researchers and research projects. The Office for National Statistics is one of the data processing bodies that will make data available to researchers in a safe way.

How did we carry out this review?

1.5. The questions we set out to ask of the statistics system, and of ourselves as its regulator, were:



1.6. The review covered all aspects involved in bringing together data, from initial question setting and establishing whether there are suitable data to link, through to the analysis and ongoing curation or eventual destruction of the data. Figure 1 contains a highly simplified mapping of these various stages (the timing or sequencing of which will differ depending on the nature of the data shared, and who is requesting the data; some, such as privacy assessments, will be ongoing throughout the process rather than a discrete step). See Annex 3 for details of further data linkage guidance materials.

FIGURE 1: data sharing and linkage stages – a simplified map of the process

1.7. Between the summer of 2017 and early 2018, we spoke to people from over 30 organisations with interests in different aspects of data linkage: academics, government statistics producers and analysts, independent researchers, data experts and research funders. We summarised the common issues from those conversations into a smaller set of themes, which we have used to identify some potential actions to address existing barriers to effective data sharing and linkage in government. We shared our preliminary analysis and recommendations with review participants and senior leaders in organisations with a role or interest in bringing about change and refined our thinking based on feedback. We have

been heavily dependent on contributions from statistics users, producers and other interested parties and are very grateful for their participation (see details in Annex 2).

1.8. The next section outlines our findings. We begin by setting out our vision of the high-level outcomes the statistics system needs to achieve so that - with appropriate and robust data safeguarding provisions in place - sharing and linking data is as easily done as said. We then outline the main areas where improvement is needed and highlight the success stories we heard about that give us confidence that the overall vision is achievable. All the recommendations made in the report are brought together in the final section where we also outline our next steps for taking this forward.

A note on language

This report's title "joining-up data for better statistics" is intended to be a short, accessible description of our aspirations for the statistics system's use of data linkage – by which we mean bringing data from more than one source together securely to help provide statistics to answer questions. We use the term linked data to describe the data that are created by the data linkage process. The terms data linkage and linked data are not widely understood by people outside the statistics system. We have used these throughout the report because we expect most of our readers will be familiar with them, but it would be useful for statistics producers to agree more accessible terms for use when engaging with the public. The choice of terms should be informed by the results of consultation with members of the public.

Chapter 2: What would a safe and effective data linkage system look like?

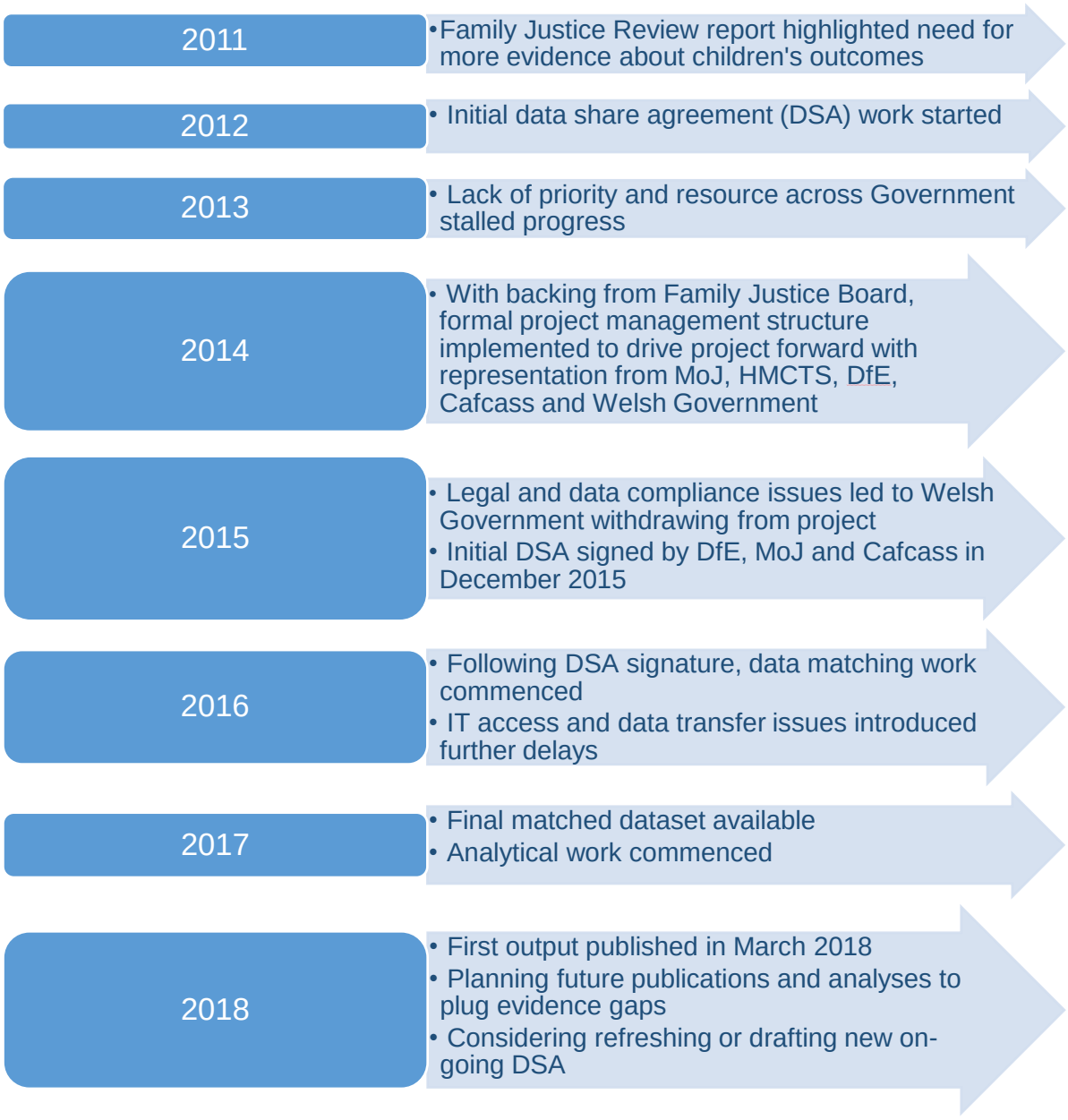
Essential outcomes

2.1. Many of our reviews looking at the public value of statistics in different thematic areas have concluded that statistics producers need to do more to bring data together from multiple sources to answer people's questions. For example, this was a finding of our recent systemic review of UK statistics on housing and planning where people want to better understand what happens, in what timeframes, from the point of land being sold to houses being built. Similarly, our justice review identified that people want to understand the various pathways people take through the criminal justice system from their engagement with the police, through the courts and potentially into sentencing. The DEA represents major progress towards realising these kinds of aspirations. However, the DEA alone cannot address all the blockages in the system that we heard about in the course of our investigation.

2.2. There is much to celebrate already. The examples highlighted in this report are by no means a comprehensive picture of all the ways that data linkage is adding value to government statistics and research. Statistics producers have for a long time been thinking creatively about ways to bring data together to help answer important questions, to inform policy development and to evaluate policies once they are implemented. However, the concern is that the efforts and time required to create these linked data resources can discourage others seeking to do similar work.

2.3. An illustration of the time that a data linkage project can take from initiation to published analyses, and the obstacles that can occur along the way, is provided below. It maps out the process that created the Children in Family Justice Data Share, involving data held by the Ministry of Justice (MoJ), Department for Education (DfE), and the Child and Family Court Advisory and Support Service (Cafcass). Welsh Government and Her Majesty's Courts and Tribunals Service (HMCTS) were also involved. Data has now been published based on the data share, but it is notable that despite the time already devoted to this, more data sharing agreements still need to be established in order for this to continue and deliver real value. The DEA could obviously help to resolve some of the legal aspects of projects like this in the future, but it will not be able to assist in all cases, nor can it address other factors that delay progress, such as resource constraints, risk averse cultures around data sharing, poor data quality and IT incompatibility.

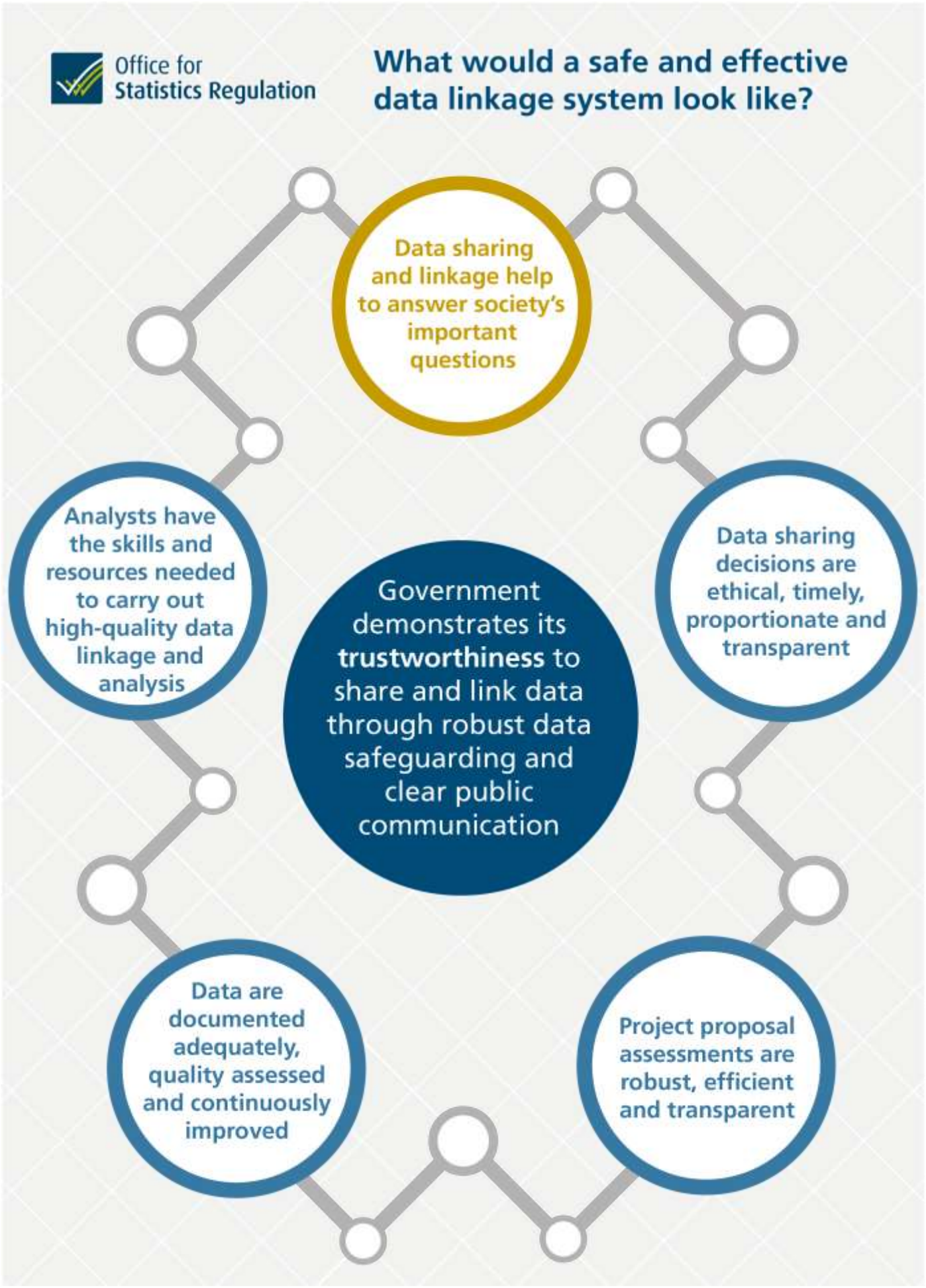
FIGURE 2: Timeline for the Children in Family Justice Data Share



2.4. Our work was also informed by the experiences of the Administrative Data Research Network (ADRN) in recent years. The ADRN was funded from October 2013-July 2018 with the aim of increasing social researchers' safe access to administrative data held by government. Its formation was recommended by the Administrative Data Taskforce, which was convened in 2011 to examine how best to make administrative data available for research safely. While our primary interest is data sharing and linkage for official statistics production, many of the challenges the ADRN faced – and its successes too – provide valuable insights for the statistics system. Access to data proved to be difficult, particularly data held by UK government departments, for many of the reasons that are outlined in this report, so its ambitions failed to be realised in full. In contrast, its methodological work, and projects using data held in Scotland, Wales and Northern Ireland, were more successful and strong links have been developed between those administrations and the researchers from the network. An announcement about UK Research and Innovation's (UKRI) plans for continuing its investment in administrative data is due soon. This investment is intended to support researchers outside government to make effective use of the new research data access provisions in the DEA.

2.5. Figure 3 sets out the outcomes we think are essential for the system to meet the ultimate objective of maximising the value of data linkage to create statistics that provide insights to users. Each outcome makes an essential contribution towards this ultimate ambition, and the ultimate ambition is dependent upon each outcome being achieved. At the core of this model, acting as a precondition for all other outcomes, is a system that understands the importance of demonstrating its trustworthiness as a custodian of public data. This demonstration of trustworthiness should serve as a platform to support greater use of data sharing and linking to deliver insights about important issues.

FIGURE 3: Essential outcomes for a safe and effective data linkage system



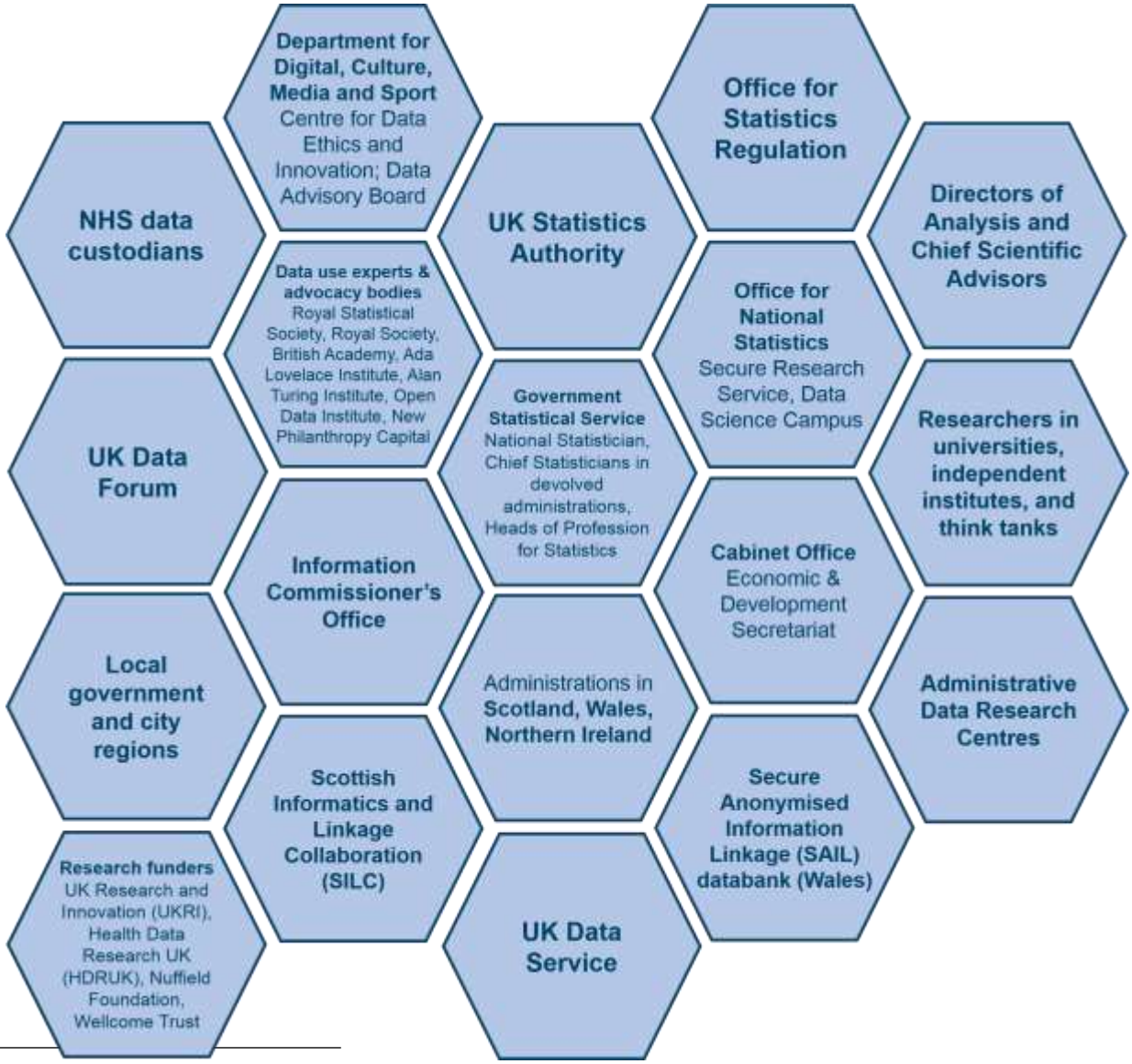
Who is responsible for delivering this?

2.6. It was clear from our review that many organisations have an interest, and part to play, in delivering the changes required to ensure that data sharing and linkage are used to maximum effect for statistics production. Examples of many of the different people and organisations with this interest are presented in Figure 4. This is not an exhaustive list and we will continue to identify other interested parties. The key to success will be harnessing and coordinating all available opportunities for leadership on data sharing and linkage issues.

2.7. As the independent regulator of the UK's official statistics system we have a clear **advocacy** role here to influence the work of the organisations in Figure 4 by using our voice to emphasise the importance of linkage for producing official statistics. We can also ensure that the importance of statistics, statistical research and the opportunities for making better use of existing data through sharing and linkage is recognised in wider debates about data ethics and the potential of other new forms of data use, such as Artificial Intelligence and big data. We can also hold statisticians to account for their trustworthiness around data governance and the quality of the statistics produced. This advocacy role is recognised by others too. As we set out in the introduction, the trustworthiness, quality and value framework set out in the new Code of Practice for Statistics underpins the **unique and useful** contribution we can make here.

2.8. The public clearly has an important role to play here as well, as both users and providers of data. Many of our recommendations are specifically intended to increase opportunities for public scrutiny of, and engagement in, major decisions about data use.

FIGURE 4: Landscape of organisations with an interest in data sharing and linkage



Demonstrating trustworthiness

What does the Code of Practice say?

2.9. Parliamentary legislation and regulatory frameworks provide the legal basis for organisations to share and link data. The Code of Practice for Statistics' trustworthiness pillar outlines what we expect of statistics producers as data custodians:

> …look after people's information securely and manage data in ways that are consistent with relevant legislation and serve the public good

This principle is supported by practices that stress the need for transparent and accountable procedures to protect personal data.

2.10. The sharing and linking of personal data needs to have a legal basis. Beyond this, organisations who share and use data should take further steps to build public acceptance in addition to relying on their legal powers – this is often described as gaining **social licence**. The concept of social licence originated in the Australian and Canadian mining industries in the 1990s and is now applied to other activities that have potential environmental impacts. More recently it has been used in the context of data in Australia, New Zealand and the UK. In this environment, social licence refers to the additional steps that need to be taken - beyond any legal and regulatory frameworks - for data sharing and linkage to have public acceptability.

2.11. With mining, social licence matters because what is being proposed often has significant public opposition. With data use, social licence matters because the legislation that allows for data sharing and linkage comes with in-built provisions around safeguarding personal data and ensuring its use is in the public interest. It can therefore be less a case of gaining social licence in the face of opposition, and more a case of **maintaining** social licence by demonstrating how these provisions are being met and that the legislation is being used appropriately. Proactively seeking to build and maintain social licence around data use should, in the long-term, also help to increase public understanding of the benefits and risks of data sharing and linking. The public will then be in a better position to engage in debates about new proposals for data use, and to judge the consequences of breaches, if they occur, either with government data or that held by private bodies.

2.12. In the UK context, the recent passing of the DEA was an important milestone. However, it should not be seen as the end of the public engagement process. The DEA's development was informed by an open policy making process facilitated by the organisation Involve with government, civil society and academia. This kind of approach should be continued to ensure that the public remain informed about and engaged in developments in data use now that the Act is in place.

What does public interest mean?

It is a condition of the Digital Economy Act (2017) that access to data will only be sought for statistics production or research projects that are in the **public interest**. The Statistics Authority's criteria to assess research projects – which could usefully be adopted by other data holding bodies – are:

The primary purpose of a research project must serve the public interest in one or more of these ways:

- provide an evidence base for public policy decision-making
- provide an evidence base for public service delivery
- provide an evidence base for decisions which are likely to significantly benefit the economy, society or quality of life of people in the UK, UK nationals or people born in the UK now living abroad;
- replicate, validate, challenge or review existing research and proposed research publications, including official statistics;
- significantly extend understanding of social or economic trends or events by improving knowledge or challenging widely accepted analyses; and/or,
- improve the quality, coverage or presentation of existing research, including official or National Statistics.

2.13. For us, social licence is fundamentally about trustworthiness and value. The steps that are needed to maintain social licence are identical to the expectations we place on statistics producers to safeguard data transparently, and to produce statistics that provide insights of value to the public.

2.14. Trustworthiness and social licence also matter because they set the limits of what can be achieved. There is strong evidence that public concerns about, and support for, data sharing is highly specific to the data and proposed uses in question (for example, work by the Royal Society in 2017). This is evident in the different conditions for accessing health versus other types of data held by government, and in the different conditions for accessing health data in the different parts of the UK. We spoke to users in government, universities, research institutes and think tanks who are frustrated by what they feel are unnecessarily prohibitive restrictions on access to NHS data in England for research purposes. Compounding this, health and social care data are excluded from the research strand of the DEA and many of the people we spoke to during this review expressed concern about this exclusion. Equally, we know that NHS Digital is not, at present, confident that it has the social licence to share data in the ways that users would ideally like to see happen. In contrast, NHS data controllers in Scotland and Wales view their social licence to use data to support research rather differently. They have less restrictive - though highly regulated and safeguarded – systems in place that have been developed over the years in tandem with various public engagement activities (see box), all of which help to maintain their social licence to operate.

Examples of actively seeking input from the public

The Administrative Data Research Network's work was supported by a programme of public engagement which included panels of members of the public to scrutinise and challenge their work. For example, the panel in Scotland debated the question of research data retention and whether organisations should move away from destroying data at the end of projects and instead curate it so other researchers can make use of it (with appropriate safeguarding in place). In Wales, the members of the public panel that supports the work of the Secure Anonymised Information Linkage (SAIL) databank play a number of roles: acting as advisors on issues in research, advising on how best to engage with the public, offering guidance on how to recruit people to study steering groups, providing views on data protection issues, discussing proposals for research, reviewing information designed for a lay audience, and acting as advocates for data linkage research. Public engagement involves much more than public panels - additional mechanisms are needed to ensure that all community voices have the potential to be heard – but these examples illustrate the valuable role the public can have.

Communication is key

2.15. We heard many examples of good practice around data safeguarding in the course of the review demonstrating that statistics producers and data holders take their duties here very seriously. However, to demonstrate trustworthiness statistics producers must do more than just ensure that robust safeguarding provisions are in place and are followed. They must also be transparent about their safeguarding procedures so the public can scrutinise the steps that are taken.

2.16. Statistics users demand consistency and coherence when multiple organisations produce statistics on similar topics, and the Code of Practice requires producers to address this. The statistics system should also ensure that information about data safeguarding is consistent, coherent and accessible. ONS has developed an excellent framework, the "Five Safes" (see box), that summarises the steps that must be taken to safeguard data. Official statistics producers across the UK should use this framework when describing their approach to data safeguarding. Over time, statements about keeping data safe should be as prominent and commonplace as statements about data quality. This framework will need to be updated as the data environment evolves and as more bodies use it and identify ways to enhance its value.

Use consistent and coherent messages: The Five Safes

Safe **data**: what steps have been taken to remove items that could identify individuals?

Safe **people**: do the users have the necessary technical skills to use the data, do they understand the importance of data confidentiality and have they completed all necessary training?

Safe **projects**: is it an appropriate use of the data, ethical and clearly for the benefit of the public?

Safe **settings**: where will the data be used and what steps are in place to ensure the data are kept safe?

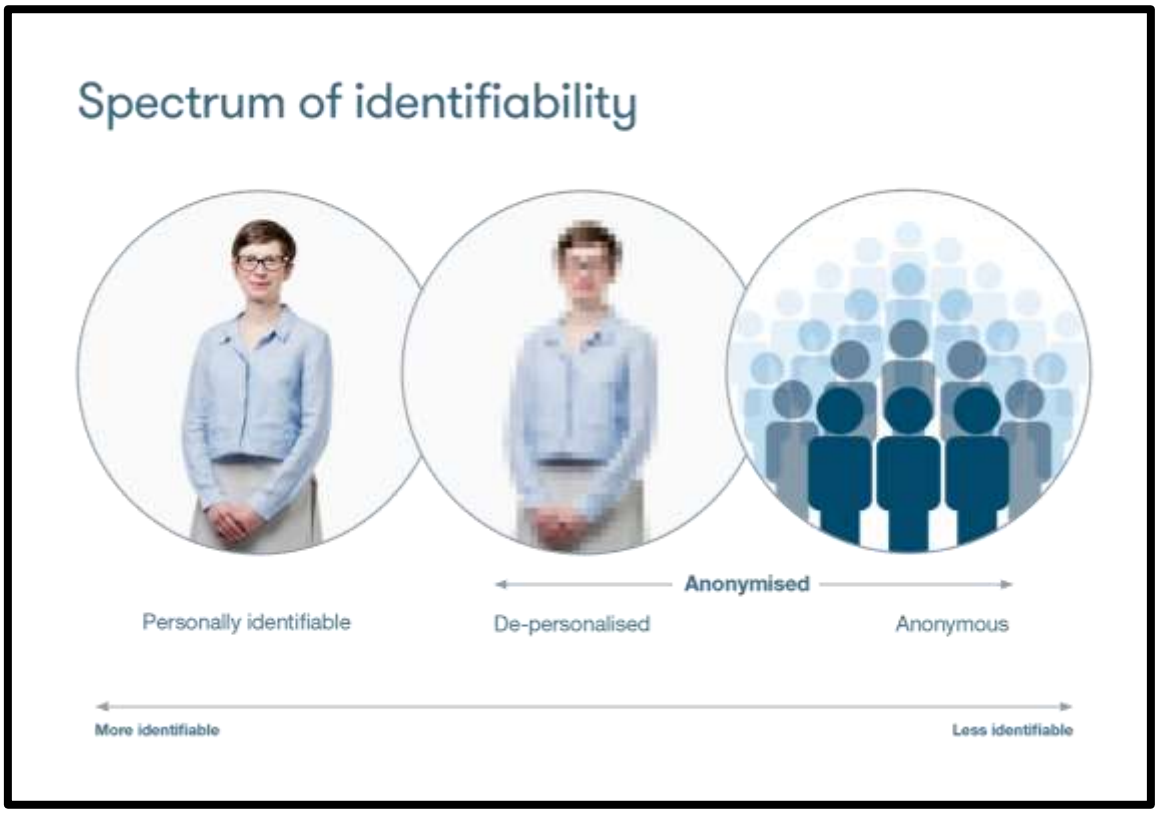
Safe **outputs**: what processes are in place to ensure that outputs produced from the data cannot be used to identify individuals?

2.17. Data safeguarding is an area with lots of highly technical language – for example 'pseudonymised', 'reidentification' and 'encryption' - and care must be taken to ensure that the information presented under each heading is meaningful for the public. Terms such as pseudonymised, reidentification and encryption are not generally helpful. In the Introduction we outlined our view that data linkage is probably not a helpful term either. Advice exists to support better communication with the public (see box below), and this is an area where the Government Statistical Service (GSS) should support analysts with guidance

Use clear and meaningful language and visual tools

The Wellcome Trust's Understanding Patient Data programme has developed guidance about communicating complex issues such as data anonymity and identification risk. Although developed with a focus on health data, the advice is relevant to all government statisticians. Their spectrum of identifiability uses helpful freely available images to explain identification risk. The Department for Education has recently drawn on these resources to develop new methods to assess and communicate identification risk, which will lead to clearer public communication around data sharing.

.



'Identifiability spectrum' by Understanding Patient Data is licensed under CC BY.

Coherence matters

2.18. The legislative and regulatory frameworks for using data to deliver services and using data for statistics and research differ, and do so for good reason. However, public understanding of these distinctions is arguably less clear cut and there needs to be a coherent approach to communication around all government uses of data. Increasingly there are also moves to use data analysis techniques to target services, either at specific individuals, or particular groups, that start to blur the lines between data for service provision and data for statistical purposes. It is also the case that some data shares for operational purposes have the potential to impact on public perceptions of data collection and its governance and security. The data shares for immigration enforcement between the Home Office and NHS Digital (currently suspended) and DfE are recent examples of this. These situations can affect trust in, and public acceptability of, statistical uses of data, with wide ranging consequences including individuals declining to participate in data collections and research data access being curtailed.

2.19. It is also important to recognise that many conversations about data use are taking place in specific localities, for example in areas where local public service agencies are testing their local populations' appetites for increased data sharing to improve service delivery. Involve is conducting work on data sharing by local authorities, supported by the Carnegie Trust UK and the Wellcome Trust.

2.20. New Zealand recognised this need for a cross-sector, coherent, approach to data use by establishing an independent Ministerial advisory group in 2015, the Data Futures Partnership, to advocate for the safe use of data to unlock economic and social value (this followed on from work conducted by the New Zealand Data Futures Forum in 2014). Following a large public engagement programme, they developed guidelines for building social licence based around eight questions that need to be answered for the public to feel confident that their data are being used safely and appropriately (see box).

Guidelines for trusted data use: example from New Zealand's Data Futures Partnership

2.21. The trusted data dial provides an overarching – and increasingly recognisable – framework for the public to scrutinise organisations' use of their data in all applicable settings (e.g. to deliver services or to produce statistics). The Five Safes framework is then used to explain Statistics New Zealand's approach to data safeguarding for specific data resources, such as their Integrated Data Infrastructure which links together data from multiple sources. The Department for Digital, Media, Culture and Sport (DCMS) in England is currently developing its new government data strategy and should consider developing a similar overarching framework for building trust in data use.

2.22. These examples from Scotland, Wales and New Zealand share a common advantage: they support the work of statistics systems with a relatively small number of highly-coordinated organisations that hold and use public data. The wider UK context is more fragmented. There will therefore be greater challenges to develop similar approaches to building social licence (for statistics or data use more generally). Additionally, the responsibility for building social licence for data sharing and use extends far beyond government (see the health data example below). We will facilitate conversations between the multiple sets of data holders and other parties across government and beyond who have an

interest in this area to help encourage more understanding and sharing of best practice around social licence issues.

Explain the benefits of using data and how it is safeguarded - #DataSavesLives

The Wellcome Trust's #DataSavesLives campaign maintains a list of case studies to illustrate different uses of NHS patient data. To coincide with the launch of the national data opt-out programme in England in May 2018, animations and films based on these case studies were posted on YouTube and promoted via social media channels.

Answering society's important questions

What does the Code of Practice say?

2.23. The Code of Practice's value pillar sets out our expectations for statistics and data that are:

> …useful, easy to access, remain relevant, and support understanding of important issues.

There is an efficiency aspect as well (V5.1), highlighting the importance of effective data use and avoiding duplication of effort:

> Opportunities for data sharing, data linkage, cross-analysis of sources, and the reuse of data should be taken wherever feasible.

2.24. Many of society's most pressing questions cannot be answered using single data sources. Data sharing and linking are therefore fundamental to ensuring that statistics deliver value. For example, OSR has been actively encouraging producers to bring together data across government to answer society's important questions about housing, migration and the justice system. There is a social licence aspect here too: demonstrating how the value of statistics is increased through safe data sharing and linking provides reassurance that these activities meet the public benefit requirements set out in the legislation.

2.25. Two ingredients are needed to answer society's important questions: effective mechanisms to **identify** questions and effective mechanisms to **answer** them. Government analysts and policy makers should do more to identify the questions that society wants to answer and to make use of the considerable resources that are being invested to support the statistics system's ambitions for better statistics and better decisions.

Mechanisms to identify questions

2.26. There are many great examples of successful data linkage work happening in government, or on its behalf, to improve official statistics and answer important questions. We highlight a small selection below, each of which required data to be brought together from more than one source because no single dataset could provide these answers. The joined-up thinking required to drive initiatives such as these should be the norm, but is still the exception. It is also the case that data linkage is currently more commonly used to support statistical research than for official statistics production. The new government analytical function should help to facilitate a more joined-up approach and create

opportunities for statisticians, social researchers and economists to learn from each other's use of data linkage.

Success stories: answering important questions…

Is Scotland's Out-of-Hospital Cardiac Arrest Strategy helping to save lives?

This strategy aims to increase survival rates, potentially saving an additional 1,000 lives by 2020. Linked data from eight sources held by the NHS, National Records of Scotland and Scottish Government is being used to evaluate the strategy's impact and inform decisions about the most effective locations for community defibrillators.

How do students benefit from going to university?

The Longitudinal Education Outcomes dataset links five different sources of data held by the Department for Education, Department for Work and Pensions and Her Majesty's Customs and Revenue. Statistics are being produced to illustrate the long-term career and earning prospects for students graduating from undergraduate courses in the UK.

What kinds of interventions can help to increase prisoners' employment prospects and reduce reoffending on release from prison?

The Justice Data Lab can match data on prisoners who have taken part in interventions during their time in prison to Police National Computer data to assess their impact on reoffending rates. A data share between the Ministry of Justice, the Department for Work and Pensions, and Her Majesty's Revenue and Customs can now be used to look at other outcomes, such as unemployment and employment, following release. For example, reoffending rates were lower, and employment rates were higher, among recipients of distance learning grants provided by the Prisoners' Education Trust than was the case for a comparison group selected to match the characteristics of people receiving grants.

What are the characteristics of people living beside Northern Ireland's Peace Walls?

The Northern Ireland (NI) Executive has committed to remove all physical interface barriers between communities (also known as 'Peace Walls') by 2023. Academics working in collaboration with the NI Department of Justice (DoJ) used linked health, mortality and deprivation data to produce baseline estimates of the characteristics of people living beside Peace Walls. These analyses are informing the DoJ and Interfaces Programme Board's work in support of the 2023 removal target

How do university student suicide rates in England and Wales compare with the general population?

Concern about student mental health led to ONS bringing together mortality data with information from the Higher Education Statistics Agency to produce experimental statistics about suicide rates in university students. Without data linkage this important issue could not be analysed. Between the 12 months ending July 2013 and the 12 months ending July 2016, higher education students in England and Wales had a significantly lower suicide rate compared with the general population of similar ages, though there is some evidence that rates might have risen in recent years (small numbers make trends hard to establish).

What happens to children in contact with Family Courts in England and Wales?

The Children in Family Justice Data Share links child-level data in England linked from across the Ministry of Justice, the Department for Education and the Children and Family Court Advisory and Support Service. It is designed to increase understanding of how decisions in the family court impact on children's educational outcomes and life-chances. The Nuffield Foundation's new Family Justice Observatory is creating a Data Platform and Analytics Service that will support an extension of this work to allow outcomes for children in Wales to be analysed for the first time using Cafcass Cymru and Welsh Government data.

Does employment support delivered in health settings improve health and employment outcomes for people with health conditions?

The Department for Work and Pensions and Department of Health and Social Care have formed a joint Work and Health Unit to lead the work across government to improve employment outcomes for disabled people set out in Improving Lives: the work, health and disability green paper. The Work and Health Unit, with NHS England, are working with West Midlands and Sheffield City Region Combined Authorities to test whether Individual Placement and Support can improve health and employment outcomes for people with physical and/or mild to moderate mental health conditions in new health settings. The Trials will use management information, survey data, and linked administrative data to evaluate outcomes. The trial data will be available for assessed and approved secondary analyses.

Do home energy efficiency measures in Wales improve health and wellbeing?

As part of its strategy to reduce fuel poverty in Wales, the Welsh Government implemented a demand-led fuel poverty scheme called Warm Homes Nest to improve the energy efficiency of homes. Administrative data for the Nest scheme was anonymously and securely linked to routine health records. The analysis demonstrated a statistically significant positive effect on the respiratory health of recipients, with significantly fewer respiratory GP events. The study led to a change in the eligibility criteria for the Nest successor scheme with eligibility extended to low income people with respiratory and circulatory conditions. The relationship between energy efficiency improvements and mental health is now being investigated with a view to further informing policy decision-making.

2.27. Some of these examples are based on work supported by the ADRN's centres located in each of the UK's four countries. Its website has many examples of projects with a strong policy focus, particularly those based in Scotland, Wales and Northern Ireland, where analysts, with strong support from Ministers and senior officials in the devolved governments, worked collaboratively with their local centres to develop projects that could inform their work. In contrast, many of the ADRN's project proposals involving data held by UK-wide or England-only departments were led by academics without this direct input from government analysts.

2.28. People we spoke to across the UK, inside and outside of government, contrasted the opportunities for linking data held by the administrations in Scotland, Wales and Northern Ireland, with the more limited access often granted to data held by Westminster-based departments. This is a systemic issue and is only partly explained by the more streamlined governance processes that are possible in smaller administrations (helped, for example, by the fact that many transfers of key data happen within one legal entity). People talked about the relative ease with which key decision makers, practitioners and analysts can be brought together in Scotland, Wales and Northern Ireland to identify and agree priorities for analysis (see example below). It also helps that departments in Scotland and Wales have been replaced by broader directorates, which breakdown barriers between policy areas. In contrast, Westminster departments are larger, more fragmented, and opportunities to collaborate and generate ideas require more effort to make happen.

Bring together key policy makers, analysts and external experts

In 2013, the Northern Ireland Statistics and Research Agency (NISRA) established the Administrative Data Forum (ADF), consisting of representatives from NI government departments, NISRA staff and senior researchers from universities. Its purpose was to identify administrative datasets from across government and prioritise work streams based on the requirements of academic and government researchers, balanced against policy need. This collaborative approach led to the development of a data acquisition strategy with a priority focus on education, benefits and health data.

2.29. Opportunities to facilitate conversations between departments (including the devolved administrations) already exist, but they could be used more constructively. Departmental Heads of Profession for Statistics, Directors of Analysis, the National Statistician and Chief Statisticians of the devolved administrations, and departmental Chief Scientists could all provide useful leadership here. All these people have regular meetings that could dedicate more time at regular intervals for identifying and prioritising opportunities for addressing cross-cutting questions. The Cabinet Office's Economic and Development Secretariat provides a useful function by identifying cross-cutting agendas, for example the current work on loneliness. DCMS's wider drive to see better use of data could also be helpful here by supporting the development of a culture that sees data sharing and linking for analysis as part of the core mission of government.

2.30. Fostering more collaborative, cross-departmental thinking about data linkage opportunities also requires closer working between analysts from different professions. The 2015 Nurse Review recommended UK Government departments take a more strategic approach to their research and development and they are now producing Areas of Research Interest statements (15 published so far). They vary in their detail and approach, but there is limited evidence of departments working either collaboratively or cross-professionally on these statements to identify cross-cutting areas of research interest (the same interests often appear in multiple statements). These statements are an opportunity for more joined-up thinking between researchers, statisticians and other analysts within and across departments.

2.31. There is a role for external input to help shape questions that government data could answer. UKRI's strategy for its future investment in admin data will need to identify priority areas for new data analysis. The ADRN developed a new model of working in 2016/17, to bring together government analysts, policy officials, academics, think tanks and third sector interests to identify common research interests around particular themes. These covered: children, world of work, ageing, productive society, and health and society (the final theme has been slower to develop because health and social care data are not covered by the DEA's research provisions). ONS and the devolved administrations are now working to curate multi-purpose linked data resources, using the research provisions in the DEA, based on the needs identified in this process. This model of working can support the statistics system to meet its obligations to users under the Code of Practice's value pillar and we will take an active interest in seeing how it progresses. Initiatives like this could also be used to help set priorities for investment as part of the next spending review.

2.32. The steps identified above demand better communication between people to bring about more joined-up working within government. Synthetic datasets could also provide a technical solution that might help with the scoping of potential analyses. Lots of people we spoke to highlighted the fact that data governance arrangements specify that projects need tightly-defined research questions (to meet the 'safe projects' requirement in the Five Safes). In contrast, Sweden's approach supports broader work programmes, so data can be used to formulate and refine research questions before they are finalised. Provisions in the UK's Data Protection Act around proportionality also mean that access will sometimes only be granted to subsets of data, rather than complete datasets. However, prior access to data is sometimes necessary to estimate sample size requirements accurately. Synthetic datasets might help address both these issues by allowing analysts to refine questions and estimate sample size needs before applying for the data. Synthetic datasets have the properties of "real" datasets but contain artificial individual-level cases to reduce – but not necessarily eliminate – the risk of disclosure. The private sector often uses synthetic data when commercially confidential data cannot be shared. Synthetic data creation is complex and may not be the right solution in all cases, but its potential contribution here should be explored. The National Statistician's Quality Review (NSQR) of privacy and

data confidentiality methods currently underway is exploring the potential for synthetic data. Other important collaborations such as the UK Anonymisation Network have an interest here.

Making ethical, timely, proportionate and transparent decisions about data sharing

What does the Code of Practice say?

2.33. The first practice (T6.1) in the Code of Practice's data governance principle stipulates:

> All statutory obligations governing the collection of data, confidentiality, data sharing, data linking and release should be followed. Relevant nationally – and internationally – endorsed guidelines should be considered as appropriate. Transparent data management arrangements should be established and relevant ethics standards met.

2.34. Adherence to this is crucial. However, we are aware that meeting these important obligations can often involve lengthy processes, particularly risk assessments of proposed data shares, that can delay projects by months, if not years. It is therefore important to find the right balance between ensuring data sharing agreements are given the thorough risk assessment they require and meeting the needs of data users in a timely way. Transparency around these processes supports the overarching objective of the Code of Practice's trustworthiness pillar, enabling both the people who have requested the data share, and the wider public, to understand what decisions have been taken and why.

Ethical data sharing and linkage

2.35. As the legal and technological possibilities to use data expand, the ethical scrutiny of data use becomes increasingly important, as reflected in the establishment in 2018 of the Centre for Data Ethics and Innovation under the sponsorship of DCMS, and the independent Ada Lovelace Institute. Proposals to share and link data for statistics and research require robust and transparent ethical assessment by independent, appropriately skilled people. In 2015, the UK Statistics Authority established the National Statistician's Data Ethics Advisory Committee (NSDEC) to provide ethical oversight of the Authority's work to improve statistics through the better use of data. The majority of NSDEC's members are independent of the Authority, providing valuable external challenge where necessary, and all decisions are published to ensure that the process is transparent and open to scrutiny. It advises the National Statistician about whether the access, use and sharing of public data for research and statistical purposes is ethical and for the public good. It considers projects and policy proposals from ONS, the wider Government Statistical Service (GSS) and the third sector.

National Statistician's Data Ethics Advisory Committee: Ethical principles

- The use of data has clear benefits for users and serves the public good.
- The data subject's identity (whether person or organisation) is protected, information is kept confidential and secure, and the issue of consent is considered appropriately.
- The risks and limits of new technologies are considered and there is sufficient human oversight so that methods employed are consistent with recognised standards of integrity and quality.
- Data used and methods employed are consistent with legal requirements such as the Data Protection Act, the Human Rights Act, the Statistics and Registration Service Act and the common law duty of confidence.
- The views of the public are considered in light of the data used and the perceived benefits of the research.
- The access, use and sharing of data is transparent, and is communicated clearly and accessibly to the public.

2.36. The DEA requires that any proposed uses of data for research purposes are ethical and NSDEC will provide ethical scrutiny of projects if required. A self-assessment process has been developed for analysts to help them build ethical considerations into their projects from the outset, it can also be used to help decide whether a full review by NSDEC would be beneficial. Training materials are currently being developed to increase understanding of the NSDEC ethical principles (see box) and how to apply them; these will need to be available to all potential NSDEC users.

Effective management of risk

2.37. Data linkage often requires decisions to be taken by people in multiple departments and organisations, who will have different sets of issues to consider and, perhaps most critically, different processes to go through to make, approve and document their decisions. Central government is not alone in this; data sharing in local government and the NHS faces many of the same challenges and sometimes additional ones. There are good examples (see box) of developments that could help central government departments to achieve more streamlined processes for data sharing decision-making. For example, by encouraging departments to agree a common approach to information governance and ensuring sufficiently senior people are providing visible support for such initiatives.

Streamline processes and provide strong leadership
The Greater Manchester Combined Authority has a clear overall vision for using data to improve the lives of people in Greater Manchester. This vision is supported by strong leadership and support from the Mayor and Chief Executives of the bodies that comprise the Authority. They are using the Information Sharing Gateway developed by local government and NHS organisations in north-west England to streamline the risk assessment, administration and documentation of data sharing in the public sector. Using this ensures that all organisations involved follow a common information governance approach, which is fully aligned with the Data Protection Act's requirements.

2.38. Efficient processes matter, but the high level of risk carried by individual data controllers is also an issue. We heard how this creates a culture of caution that affects all decisions taken. These risks are broad and cover concerns that apply universally to all data (such as misuse and reidentification), but also concerns that are unique to data for statistical or research purposes, such as the potential for resulting analyses to be politically unwelcome (discussed further in the next section). These risks can have very direct consequences – reidentification has become a criminal offence in the Data Protection Act 2018 – so data controllers are understandably cautious. When people move posts, their replacements inherit these risks, so a reassessment often happens (if the process is at a stage that would allow it), resulting in further delays. We also heard about applications held up because new approval processes were introduced after their initial application had been submitted and processed, requiring them to start again from the beginning. These situations will sometimes be unavoidable and can be essential for data safeguarding, but data holders must ensure that delays are minimised, that new processes genuinely add value, and must communicate effectively with users about why they are happening.

2.39. One senior data controller we spoke to suggested developing a tool to help assess reidentification risk. This would be used to aid decision-making by helping to set the conditions under which the data could be accessed, or help set the threshold above which projects would be rejected. This would enable researchers to refine their requirements until their data request met the acceptable level of risk. Objective frameworks to manage risk could help reduce the personal burden felt by data controllers. The NSQR of privacy and confidentiality methods is covering techniques for quantifying disclosure risk and information loss. Open Data Institute, Alan Turing Institute and ONS's Data Science Campus have useful expertise here.

2.40. Other people we spoke to noted that, in many cases, data requests will be made by organisations or individuals with an established track record of safe and appropriate use of data. There may well be scope for the risk assessment process to therefore take account of this kind of information to develop a clearer trust model that factors in prior experience. The Open Data Institute is currently investigating the potential for data trusts to help increase access to data while maintaining public trust. Data trusts were recommended by the 2017 independent review into the growth of artificial intelligence in the UK as a means to 'share data in a fair, safe and equitable way'. As regulators of the statistics system, OSR

has a close interest in innovative approaches to data governance and will monitor developments like these with interest.

2.41. We also heard that the 2018 Data Protection Act (DPA), which implemented the General Data Protection Regulation (GDPR), has left some data holders with questions about their new responsibilities under the Act. The UK Statistics Authority has worked with the Information Commissioner's Office (ICO) to develop new guidance about GDPR and its application to statistics production and statistical research, including data sharing and linkage. This will be published later in 2018 and should help to address these concerns.

2.42. Traditionally, many data linkage projects have been required to destroy any data once the analysis is complete – known as a 'create and destroy' model. This is increasingly seen as a wasteful way to treat data which, providing appropriate safeguards are in place, could continue to yield useful insights if retained for use by other approved researchers. The ADRN has developed a helpful policy on data re-use. Much of the time spent creating linked data resources is spent negotiating the approvals for the initial data share and linkage, replicating this process later if a dataset has been destroyed is not compatible with the Code of Practice's efficiency and proportionality principles. Much value will also be lost by destroying data in this way. Where appropriate, we support the statistics system developing a properly safeguarded 'create and curate' model instead. This is another area where the NSQR's work on privacy and confidentiality methods will be valuable.

2.43. These kinds of solutions will only go so far. Ministers and senior leaders within the civil service clearly have high aspirations for making better use of data to develop policy and improve services. Departments should review their information governance procedures and involve more senior people directly in the decision-making processes for data sharing, increasing the seniority of the staff expected to ultimately bear these risks – with appropriate resources to support these decisions – if necessary.

Transparency

2.44. Transparency around the risk assessment process helps to demonstrate trustworthiness. Departments can do this by conducting and publishing Data Protection Impact Assessments (DPIA), which were introduced in the 2018 DPA as a requirement of GDPR. They are mandatory where data are combined from multiple sources, so they are highly relevant for data linkage, but the ICO recommends they are also conducted on a voluntary basis for any large-scale processing of personal data. The accountability principle in the DPA requires organisations to have appropriate records in place to demonstrate compliance if required. Conducting DPIAs therefore meets the DPA accountability principle, and publishing them meets the requirements for transparency in the Code of Practice for Statistics, providing they are written in plain English, or are accompanied by explanatory text to clarify more technical points.

Mainstreaming transparency

The Scottish Government's health and homelessness in Scotland project linked local authority data about homelessness between 2001 and 2016 with NHS data on hospital admissions, outpatient visits, prescriptions, drugs misuse, and National Records of Scotland information about deaths. Alongside the main analysis report they also published the original application for the data, privacy impact assessment, public benefit and privacy panel application and correspondence documenting its approval, and details of how to access the data. This approach is now standard practice for all publications based on linked data.

2.45. Another step producers can take to increase transparency is to publish details of all data share requests made and their outcomes. The Department for Education in England has been doing this in relation to ad-hoc National Pupil Data Sharing for several years. In December 2017 broadened the scope to cover all routine sharing of personal data. They have recently consulted users about further changes to make this easier to engage with and understand. NHS Digital also provide this information. Decisions reached about applications for data under the DEA's research provisions will be published by the UK Statistics Authority, and all data shares under the DEA (for operational or analytical purposes) will be included in a register. Linked to this, it is important to be transparent about the public interests served by these data shares. There is a risk that the statistics system puts all its efforts into

documenting what data has been shared, but misses opportunities to showcase what it has been shared for; data users clearly have a responsibility here too.

Assessing projects and people robustly and efficiently

What does the Code of Practice say?

2.46. Robust data safeguarding requires projects to be assessed to ensure that data are used appropriately, while analysts must demonstrate that they are suitably skilled (the Safe Projects and Safe People requirements in the Five Safes). Again, a balance must be struck between ensuring robust assessments take place and ensuring timely access to data. No one expects these processes to be easy; but nor should they be unduly prohibitive.

2.47. As outlined in the section on trustworthiness, projects must serve public interests. However, requests for data sharing are sometimes perceived to be held-up or refused because the resulting analyses might be politically unsettling for departments. The second practice (T1.2) in the Code of Practice's honesty and integrity pillar is clear that data sharing must not be impeded by such considerations:

> The collection, access, use and sharing of statistics and data should be ethical and for the public good. Those producing and releasing statistics should be free from conflicts of interest, including political and commercial pressures, that may influence the production, release and sharing of the statistics and data.

2.48. The public value pillar of the Code of Practice places expectations on statistics producers to put user needs at the centre of their work. The same approach should be adopted for the processes for accessing data. There are multiple sets of user needs to be considered. Robust assessment processes for projects and people are there to ensure that data are only used to serve public interests – meeting the public's need for assurance on this vital point. The processes for scrutinising analysts' proposals and skills need to be efficient, supported by clear, accessible information about what is required of them. User feedback should be collected about the application process and users should be included in any planned re-designs. The new application process to the UK Statistics Authority for research data via the DEA is undergoing user-testing which should continue until the system is fully established. The transparency being built into this application process is particularly welcome.

Designing processes with users in mind

2.49. The public value pillar of the Code of Practice places expectations on statistics producers to put user needs at the centre of their work. The same approach should be adopted for the processes for accessing data. There are multiple sets of user needs to be considered. Robust assessment processes for projects and people are there to ensure that data are only used to serve public interests – meeting the public's need for assurance on this vital point. The processes for scrutinising analysts' proposals and skills need to be efficient, supported by clear, accessible information about what is required of them. User feedback should be collected about the application process and users should be included in any planned re-designs. The new application process to the Statistics Authority for research data via the DEA is undergoing user-testing which should continue until the system is fully established. The transparency being built into this application process is particularly welcome.

2.50. The wider public also benefits from efficient processes to support analysts to access data: if a project is deemed to serve public interests, then those interests are best served with timely analysis and dissemination of the results. Public money is wasted if civil servants and external researchers spend large amounts of their time applying to access data, rather than analysing it.

2.51. Potential users of health data in England were particularly vocal about their difficulties accessing data, but other users told us about similarly frustrating experiences. We heard about lengthy and duplicative processes, often compounded by complex information governance structures and lengthy risk-assessment processes by data controllers, resulting in applications taking years, to progress. In some

cases, approvals took longer than projects were originally funded for. Concerns were also raised about whether the people conducting this scrutiny are sufficiently supported to make decisions when faced with complex issues. This is not just a concern for external data users. These delays affect internal government users, impeding analysts' ability to provide insights from data in a rapidly moving policy environment. Users do not necessarily expect the process to be rapid, but a degree of predictability would allow people to plan their resources. We also spoke to data providers who raised applicants' failure to complete paperwork correctly or provide the requested information. Increasing the resources available to users, and working with users to identify what support is needed, may help avoid issues like this. NHS Digital's Data Access Request Service (DARS) has recently introduced new support for users. Universities now have a named contact to help build relationships and embed NHS Digital's expertise about specific users. Guidance explaining how the DARS works has been produced (including videos). Many data providers now have dedicated staff to support users throughout the applications process, for example the ONS Secure Research Service, SAIL data bank in Wales and eDRIS in Scotland. These are vitally important roles in the statistics system.

2.52. These positive developments need to be monitored to assess whether they are meeting their aims and to identify whether best practice examples can be shared between data providers. Health data users in England raised concerns with us directly about access so we have an ongoing interest here, particularly because the research provisions of the DEA do not include health and social care data. With the launch of the national data opt-out programme in May 2018, the new Data Protection Act to implement GDPR, and NHS Digital's recent reforms to DARS, now would be a good time to take stock of user experiences and the wider terms under which NHS Digital's decisions are made, to help inform the direction of any future developments. We will therefore work with key interested parties to scope the terms of a collaborative review of NHS Digital's data sharing and release processes. The terms of the review will be developed in consultation with data users in government departments, universities and research institutes, and NHS Digital.

2.53. The fact that the DEA research provisions exclude health and social care data means that users will continue to face multiple points of entry to access publicly held data. Data providers should help users to navigate this complex landscape by signposting other resources that might be of interest. For example, the UK Statistics Authority webpages explaining the DEA research data access processes should sign-post users to NHS Digital's DARS, and to other public data likely to be of interest to users (e.g. in devolved administrations).

The importance of user-focused processes

Scotland currently has a public benefit and privacy panel for health and social care research, and one for social science research projects. There are plans to merge these to help streamline the decision-making process and improve users' experiences.

The importance of transparent processes

Applications to the UK Statistics Authority for data under the research provisions of the DEA will be transparent so it will be possible to scrutinise the time taken to reach decisions, and what those decisions were. The independent approvals panel chair will have an advocacy role to intervene in situations where the panel judges a project to have public benefit but a department is slow to make its judgement or declines to approve it.

Documenting, assessing and improving data

What does the Code of Practice say?

2.54. Most data linkage for statistical or research purposes in government involves administrative data, typically collected as part of delivering a service. Its use for analyses is therefore a secondary consideration. Users were particularly keen for the recommendations from this review to highlight the importance of data quality and the challenges of using administrative data. Ambitions for using administrative data often unravel when it becomes clear that data quality does not meet the standards required for statistics production. It can also be the case that high quality and highly valuable data are underutilised because there is not enough information available for analysts to assess quality or make sense of its properties.

2.55. All the principles that underpin the Code of Practice's quality pillar are important for analyses based on linked data:

> Suitable data sources: Statistics should be based on the most appropriate data to meet intended uses. The impact of any data limitations for use should be assessed, minimised and explained.
> Sound methods: Producers of statistics and data should use the best available methods and recognised standards, and be open about their decisions.
> Assured quality: Producers of statistics and data should explain clearly how they assure themselves that statistics and data are accurate, reliable, coherent and timely.

Why data documentation matters

2.56. The UK Statistics Authority's Quality Assurance of Administrative Data (QAAD) toolkit sets out the approach we take as regulators when assessing statistics based on administrative data. It states (on p4):

> "The Authority encourages the application of critical judgment of the underlying data from administrative systems before the data are extracted for supply into the statistical production process. As with survey data, producers need to: investigate the administrative data to identify errors, uncertainty and potential bias in the data; make efforts to understand why these errors occur and to manage or, if possible, eliminate them; and communicate to users how these could affect the statistics and their use."

2.57. However, the approach to, and extent of, documenting administrative data is highly variable. In contrast, significant resources are dedicated to making sure data collected in surveys is fully documented so that it can support secondary analyses. The UK Data Service's guidance on documenting data describes the requirements succinctly: "make data clear to understand and easy to use". The responsibility for improving data documentation lies with the data holders, but they may need input from the statisticians using the data to identify what needs to be documented, so cross-departmental and cross-professional working may be important here.

2.58. We have an interest in this because the expectations we place on statistics producers to assure the quality of the source data used in the production of statistics cannot be met if they do not have enough information about the data. There is a value angle as well. Data cannot be used effectively – or at all, in some cases – if analysts do not understand its properties and limitations, or even know that it exists. Failing to document data could result in opportunities being lost, or poor value work being conducted, because the strengths and weakness of the data are not clear.

2.59. Recognising resource constraints, it would be valuable to identify the data sources most in need of additional documentation and metadata, and to prioritise a small number of those to be addressed initially. The Data as a Service team in ONS would be best placed to lead on this, but external expertise is likely to be helpful, for example from the UK Data Service. Many external resources already exist that could support this process, for example by helping to determine a minimum set of requirements for metadata, including those developed by Eurostat and the UN Economic Commission for Europe. The FAIR data principles, developed to make data Findable, Accessible, Interoperable, and Reusable could also be useful here. As part of this process, a register of administrative data sources,

to build on existing statements of administrative sources, should also be developed over time, with unique asset numbers, to help users navigate what is available and what data have been used for more efficiently.

Supporting the quality assurance of linked data

2.60. There are at least two stages to the quality assurance of linked data: the assessment of the underlying data sources being linked, and the assessment of the linkage process itself. The common requirement for users is that both these stages are transparent and documented (see below). We welcome the work currently being led by the GSS Methodology Advisory Committee to address the issues set out in Professor David Hand's Royal Statistical Society paper Statistical challenges of administrative and transaction data, many of which are related to quality assurance. The involvement of external experts from across the UK increases the potential for this work to have far reaching impact. Seeing progress on these issues would be a major success for the statistics system. This work highlights the value that external expertise can add to the statistics system.

Document and assess the linkage process

Data linkage involves many stages, some of which are not always transparent to the final users of the data. Academics funded by the ADRN worked with ONS methodologists to develop the Guidance for Information about Linking Datasets (GUILD) guidelines, which have been peer-reviewed by international experts. These outline the information that needs to be made available about each stage in the linkage process, including an evaluation of the linkage accuracy, to improve the transparency, replicability and accuracy of data linkage, and to support assessments of the validity of resulting analyses conducted. We have asked the ONS Best Practice and Impact Team to develop a data linkage documentation and assessment framework for government analysts, based on the GUILD guidelines.

Statistical expertise can drive data quality improvements

2.61. The steps outlined above are intended to help statistics producers and users assess the strengths and limitations of statistics based on linked data. Quality assurance can also lead to improvements if information is fed back to data collectors (see box). As well as doing more to document their data, data providers should also do more to harness information to improve data collection practices. Building in these feedback loops is more difficult for established systems. However, where new management information systems are being developed by government (either as new services are developed or old ones are digitised), or changes are being made to existing systems, these have huge statistical potential so statisticians should be involved in their design and ongoing evaluation to ensure that data can be extracted for analysis or linkage with minimal fuss. Where new contracts are being awarded to external suppliers to deliver public services then statistical input should be sought at the procurement stage to ensure that data collection and sharing opportunities are addressed before contracts are signed. The DEA requires data suppliers to consult ONS if changes to existing systems that could affect their official statistics production are being proposed. The National Statistician should use his voice to advocate for the input of statistical expertise to the design of new data systems across government, and OSR will raise this issue with the team at DCMS developing the new government data strategy.

Collect, analyse and improve data

Health and social care service integration is happening across Britain. This has shone a light on the different kinds of data recording practices within each system, and has heightened the need for more standardised measures to be collected. In Scotland, analysts at the National Records of Scotland capture information about social care data quality as they process and use it and feedback issues to data providers. This information is used to make changes to data collection practices in real time and improve data quality.

Ensuring analysts have sufficient resources and skills

What does the Code of Practice say?

2.62. The Code of Practice's trustworthiness, quality and value pillars all matter here. The statistics system needs sufficient people to carry out and assure high quality data linkage and analysis, with appropriate skills, and access to the technological resources required to support their work. Without these elements in place, innovative uses of data to add value to official statistics production cannot happen.

2.63. Investment in skills is a key aspect of this, as highlighted in the professional capability practice in the Code Practice's trustworthiness pillar:

> People producing statistics should be appropriately skilled, trained and supported in their roles and professional development.

2.64. The steps outlined in the sections above to enable data to be joined up more efficiently are necessary requirements to realise the ambition of answering society's important questions. However, specialised analytical skills are also needed to ensure that the data are used to full effect. These sets of requirements reinforce each other: increasing access to data also enables more people to develop the analytical skills required to maximise value.

Money matters

2.65. UKRI's investment in administrative data, continuing its work beyond the ADRN, will provide the statistics system with resources to support data sharing, linking and access for external research purposes. While the DEA has potentially removed some of the legal barriers that were previously in place, these funds will help remove some – though not all – of the resource barriers. However, UKRI does not fund official statistics production. Departments will need to identify resources to meet new demands for data and be realistic and transparent about what can be achieved with the funds available. Opportunities to add value via data linkage should not be missed solely because funds are not in place at the beginning of projects to carry out the linking and subsequent analysis. However, the ethical and participant burden implications of this should be balanced against the prospect of securing funding. The statistics system also needs to address the funding implications of the fact that some departments hold data of interest to a wide range of other departments and organisations so end up being net contributors of data to the system. The recommendations below for developing analysts' skills will all have resource implications, largely in the form of people's time to provide and receive training.

2.66. Looking to the future, the size of some existing datasets will grow considerably and departments will need to ensure they have sufficient capacity to store and manage their data. This could develop in a piecemeal fashion, or the system could coordinate its efforts to identify options to suit multiple needs. Analysts need to work collectively across government to identify the questions that data linkage could help to answer. This approach should also be used to identify system-wide solutions to infrastructure and resource issues, such as cloud computing solutions to address storage capacity, which would not make sense to be procured on a department by department basis. Switching from paid-for software licences to open source packages releases resources, and helps transparency as more code can be shared with external users without requiring them to have specific paid-for software. There are good channels across the GSS and Government Digital Service for sharing expertise around software and these should be utilised to share information about open source packages for data linkage.

2.67. The scale of data held by government is vast but much of its management is outsourced to external contractors who charge for each data extract requested. Access costs can be prohibitively high for some projects. While existing contracts may be difficult to modify, departments should identify creative solutions to cut the costs of data extracts and must ensure new contracts are more flexible in future.

Building skills for the future

2.68. Methodological best practice and innovation is evident in many places *within* the statistics system, but more should be done to share and promote learning about data linkage across the system. This

includes broader skills and knowledge about the legislation underpinning data sharing and linkage, ethical practice, data governance and project management skills. There is a huge appetite for this. We heard ideas for a new GSS data linkage network, greater use of technology to connect people (e.g. Slack channels), a data linkathon event, videos and other resources. However, the people within the system with the capability and desire to develop these ideas and share their expertise do not always have spare capacity to do so. Leadership is therefore required to develop a vision of what an effective multi-skilled and informed workforce should look like. A strategy and resources will be needed to realise that vision.

2.69. Much of the methodological and analytical expertise in data linkage is found in universities and independent research organisations. More use could be made of these skills to ensure that up-to-date methods for linkage and analysis are used throughout the GSS, rather than in pockets as at present. In April 2018, NHS Digital published it plans for delivering linked datasets to support health and care research (linking health datasets held by the NHS). They are seeking external guidance and peer review of their linkage methodologies, and user input via a new data linkage technical advisory group. Major research funders such as UKRI, Wellcome and Nuffield are doing much to develop the UK's methodological capacity in this area. Access to data for research purposes adds value to the statistics systems– by identifying important questions to be answered, conducting analyses that government does not have the capacity or capability to do, and by training new cohorts of much-needed analysts with the specialist skills to work with linked data. The statistics system and these bodies therefore need to work together to ensure that the contributions each can make are fully realised.

2.70. Where data linkage is concerned, we are really talking about a number of systems, many of which include formal partnerships with organisations outside the Government Statistical Service (GSS). For example, the Scottish Informatics and Linkage Collaboration (SILC) and the Secure Anonymised Information Linkage (SAIL) Databank in Wales. The ADRN has created an important cross-UK network of people with common interests and significant expertise. Some of the skills and good practice that the GSS would benefit from being shared more widely are located in these organisations and networks (and many others). New models of working could be needed to ensure the value of these external contributions is recognised, and that the staff in those organisations can also benefit from initiatives to build skills. A concrete example is training and professional development opportunities for people in research support roles who oversee research applications, accredit researchers and conduct disclosure control output checks. ONS is expanding the number of people in research support and accreditation roles; opportunities for their ongoing professional development will need to be identified.

Invest and innovate in secure data access routes

2.71. The Code of Practice's Value Pillar makes clear that statistics and data should be equally available to all. The recommendations in this report are intended to deliver a step-change in the use of data linkage for official statistics production. However, it is vitally important that these new data resources are not seen as solely for the benefit of government analysts, especially where they relate to new policy developments that should be open to scrutiny via secondary analysis. Users we spoke to expressed disappointment that a number of the key linked datasets that have been developed in recent years have yet to be made available for approved researchers to do work that would be of public benefit. In some instances this is because secure access settings are required – the safe places requirement in the Five Safes framework.

2.72. There is a huge amount of impressive work now being done to improve users' experiences of the data access process while maintaining the highest standards of data security. We were struck by how keen the data holding organisations we spoke to are to keep improving processes. Safe settings can help to safeguard data, for example, holding data in a central location for users to access remotely reduces the risk that it could be shared with people who have not been granted permission to use it. For some data, even more secure access requirements are needed, so there are physical settings where people go to access data where they need to remove all personal electronic devices before entering, and cameras record them as they work. In both virtual and physical safe settings, additional statistical disclosure control checks can be carried out of results tables before they can be taken away, to ensure that nothing that could identify an individual is released. While safe settings protect data, they place limitations on users, particularly those located some distance from physical settings (there are notable gaps in some parts of the UK), and people whose personal circumstances mean that long distance travel is not possible. Virtual access platforms are available via some safe settings and others are being developed. These are important resources for the statistics and research community.

2.73. Another data access model has been developed by the Ministry of Justice Data Lab which links data held by organisations working in England and Wales to help rehabilitate offenders with data held by the justice system on reoffending to assess the impact of interventions. It was established in response to a recommendation by New Philanthropy Capital. Data are not released directly to external organisations, instead the Justice Data Lab conducts the linkage and statistical analysis on organisations' behalf and publishes details of the projects it has supported. A similar resource to support organisations working in the employment and training sector has been proposed.

2.74. Statistics producers who want to provide users with direct access to data that requires safeguarding, but do not currently have a safe setting option, should avoid creating new systems of their own and instead use one of the existing settings (see example in box). If an existing safe setting cannot meet producers' needs and a new system needs to be created – and we would expect that to be rare now – the processes for assessing people and projects should adopt those already in place elsewhere.

2.75. Ultimately, the various secure data settings that exist across the UK in the ONS, government departments, the UK Data Service, universities and NHS should adopt common operating standards so they can form a network and each operate as a single point of entry for users to access data. The new connection being created between the ONS Secure Research Service and SAIL Databank is a very welcome step in this direction.

Take advantage of existing resources for efficient and safe data access provision

The Department for Education in England is currently working with ONS to enable secure access to data via the ONS Secure Research Service (previously known as the virtual microdata lab). At present this is via physical settings, but virtual access will be available later in 2018. This means that new data will be available to users (with appropriate safeguards) without the need to create a new safe setting, along with all the application and accreditation processes that this would require. This will also limit the number of different application processes that data users have to navigate.

Chapter 3: Next steps

3.1. The main message to take from this report is that data linkage should be a vital component of the official statistics landscape. Value is being squandered because this is not currently the case. We want analysts to feel empowered to make data linkage a core part of their efforts to innovate and improve official statistics production.

3.2. New powers have been given to the statistics system in the form of the DEA; new investment from UKRI will support researchers to access data. Users' aspirations within and beyond government are high and their demand is clear: make this work. Our interest in supporting these aspirations does not end with this review. It will be embedded into all our work to enhance public confidence in the trustworthiness, quality and value of statistics produced by government.

3.3. All the recommendations made in the previous sections are summarised below. We noted at the start that a wide range of people and organisations have some part to play to realise the overall vision we have set out. The most likely audiences for the messages in this report fall broadly into four groups. The boundaries between these groups are fluid, but identifying their different characteristics and interests in this way helped us to frame our thinking. They are:

- The **senior leaders** in government and elsewhere who enable or commission data sharing and linking. This group is chiefly responsible for making the changes required to deliver the six outcomes captured in Figure 3.
- The **data practitioners** in the statistics system (in government, arms' length bodies and the NHS) who safeguard data, carry out linkage and conduct analyses to produce statistics and statistical research.
- The **data users** outside government who provide vital additional input as analysts and methodological experts.
- The **observers and debaters** of issues about data use and ethics, in government and beyond.

1. Actively seek input from the public in major decisions about statistics production and statistical research using data linkage.
2. Identify clear, consistent and meaningful language to use when engaging with the public about data safeguarding, linkage and use.
3. Be advocates for safe data use to provide insights that serve the public interest.
4. Produce "keeping data safe" statements using the Five Safes Framework.
5. The Department for Digital, Culture, Media and Sport should develop an overarching framework for trustworthy government data use.

6. Maximise opportunities to identify the questions that society wants answered by exploiting existing networks of senior leaders e.g. Heads of Profession, Chief Statisticians, Directors of Analysis and Chief Scientists.
7. Ensure that policy makers and external experts are actively involved in processes to identify questions.
8. Ensure that departmental Areas of Research Interest Statements explore cross-cutting interests across departments, and are produced with input from all analytical professions.
9. Enable more exploratory analysis to take place before research questions are finalised, including through the use of synthetic data.

10. Agree common information governance frameworks to harmonise practice across government departments.
11. Consider placing the responsibility for signing off data sharing agreements with more senior staff.
12. Explore the contribution that risk assessment tools could make to decision-making about data shares.
13. Publish materials related to data shares, including mandatory and voluntary Data Protection Impact Assessments, to support transparency.

14. Design data access application processes and support materials with user input and seek ongoing feedback on systems when they are live.
15. Signpost users to other sources of data of potential interest to them.
16. Work with health data users and NHS Digital to scope the terms of an independent review of NHS Digital's health data sharing and access processes.

17. Identify data sources most in need of additional documentation and metadata and work with data owners to address gaps.
18. Create a central registry of administrative data sources.
19. Ensure that each step in the data linkage process is documented using a common framework.
20. Ensure that statistical experts are consulted and participate in the design process for new data systems.
21. Identify mechanisms for data users to feedback information about data quality to data collectors to help improve data at source.

22. Recognise resource needs – including the imbalance of demands placed on data holding departments – and either address them with additional inputs, or be clear about what constraints are faced and their implications.

23. Government departments to work together to identify resource-efficient solutions to infrastructure requirements (e.g. data storage space, software).

24. Identify creative solutions to cut the cost of data extracts charged by external contractors; address this issue in any future contracts.

25. Develop a new data linkage skills strategy to support the expansion of opportunities for training and development in this area.

26. Identify effective mechanisms to bring in external expertise on data linkage methods and analysis from academics and other experts.

27. Ensure the professional development needs of staff who support the data access process are met, including opportunities to network and share practice with people in other organisations in similar roles.

28. Continue to innovate and share practice around the delivery of safe data settings, especially virtual solutions.

29. Develop a network of accredited safe settings with common operating standards to act as a single entry point for data users.

3.4. Detailed actions to implement these recommendations will be discussed and agreed in collaboration with interested organisations. We expect the statistics system to treat these recommendations as a whole, and for the National Statistician to respond with an action plan. However, we recognise that the responsibility for delivering some of the specific points lies with other parties, including OSR. We therefore expect to play a part in facilitating that response and helping to deliver the overall vision in the following ways:

- We will convene meetings with key stakeholders to discuss public engagement, social licence and overarching frameworks around data use.
- The Five Safes and DPIA recommendations will be included in the new Code of Practice for Statistics data governance guidance.
- We will reinforce these messages in meetings with Directors of Analysis, Heads of Profession for Statistics, and Chief Statisticians in devolved administrations.
- We will highlight relevant recommendations as part of our engagement with DCMS over the new government data strategy.
- We will seek out and promote examples of good practice in this area in our regulatory work.
- We will use our voice to challenge statistics producers to make more use of data linkage to answer society's questions.
- We will take a more active role in monitoring user experiences of safe setting provision and data access application processes, including the introduction of the new DEA research provisions.
- We will work with health data users and NHS Digital to scope the terms of an independent review of NHS Digital's health data sharing and access processes.
- We will feed these recommendations into the upcoming refresh of the UK Statistics Authority's Better Statistics, Better Decisions strategy.
- We will work with UKRI to support the development of their strategic direction around investment in administrative data.

Annex 1: Glossary

Term	Definition
Administrative data	Data collected primarily for administrative reasons (not research). This type of data is collected by government departments and other organisations for registration, transactions and record-keeping, usually when delivering a service. Administrative data are often used for operational purposes and their statistical use is secondary.
Administrative Data Research Network (ADRN)	The Administrative Data Research Network (ADRN) was an Economic and Social Research Council-funded project that ran from October 2013 to July 2018. It brought together experts from all over the UK to organise an infrastructure to enable trained economic and social researchers to access linked, de-identified administrative data in a safe setting. Administrative Data Research Centres were based in each country in the UK. Administrative data research will be taken forward in a new ESRC-funded project to be launched in late 2018.
Areas of Research Interest statements	Statements that set out the strategic research and evidence priorities of UK government departments. Each statement should provide details on departmental research systems, research and data publication policies, and research and development strategies. The statements are a response to the 2015 Nurse review of the UK Research Councils.
Code of Practice for Statistics	The Code provides producers of official statistics with the detailed practices they should commit to when producing and releasing official statements. It plays an essential role in ensuring that the public has confidence in the statistics published by government. The framework for the Code is based on three pillars – Trustworthiness, Quality and Value – which are helpful to anyone producing data, statistics and analysis.
Data controller	The individual or legal person who controls and is responsible for the keeping and use of personal data. They determine the purposes for which, and the means by which, the personal data are processed.
Data linkage	A process which brings together two or more sets of administrative or survey data to produce information (linked data) that can be used for research or statistical purposes.
Data protection impact assessment (DPIA)	A tool or process that helps organisations identify, assess and mitigate or minimise privacy risks with data processing activities. DPIAs were introduced in the 2018 Data Protection Act as a requirement for GDPR, and are mandatory where data are combined from multiple sources.
Digital Economy Act	The Digital Economy Act 2017 (DEA) is an Act of Parliament addressing key issues relating to electronic communications services (the digital economy). The DEA amends the Statistics and Registration Service Act 2007 to support the UK Statistics Authority's statistical production function (carried out by the Office for National Statistics) with greater and easier access to a range of data sources held within the public and private sectors.
eDRIS	The electronic Data Research and Innovation Service (eDRIS) is part of Information Services Division (ISD), Scotland. It provides a single point of contact to assist researchers in study design, approvals and access to NHS Scotland data in a secure environment.
Five Safes	A framework developed by ONS that summarises the steps that must be taken to safeguard data. The Five Safes are: safe data, safe people, safe projects, safe settings, and safe outputs.
General Data Protection Regulation (GDPR)	The General Data Protection Regulation (GDPR) is a regulation in EU law on data protection and privacy for all individuals within the EU and EEA. It came into force on 25 May 2018. The UK has implemented a new 2018 Data Protection Act which largely includes all the provisions of the GDPR.

Government Statistical Service (GSS)	The GSS is a cross-government network, led by the National Statistician, spread across a range of public bodies, including components of the devolved administrations and UK government departments. The GSS community works together to provide the statistical evidence base required by decision-makers, publishing around 2,000 sets of statistics each year, and providing professional advice and analysis to decision-makers.
Legal gateway	A means by which data and statistics can be disclosed or received for particular purposes. Legal gateways may be permissive (creating a discretionary power to disclose or receive data) or mandatory (requiring data to be transferred in certain circumstances).
Linked data	The product of bringing together data from multiple sources.
Metadata	Information or data that defines and describes other data. This can be to help with the discovery and identification of data, for example, through naming and labelling; by describing different data types, relationships with other data and their characteristics; or to help with data management by indicating when and how it was created, different file types or any other technical information.
National Statistics	Official statistics assessed as fully compliant with the Code of Practice for Statistics are given National Statistics status by the Office for Statistics Regulation, in line with the Statistics and Registration Service Act 2007.
National Statistician Quality Review (NSQR)	NSQRs cover thematic topics of national importance, conducted on behalf of and for the GSS. They are designed to ensure that methodologies used by GSS to produce statistics keep pace with changing data sources and technologies. Their aim is to support the GSS in identifying what good practice looks like for these methodologies and help identify opportunities for further development and investment.
Official statistics	Statistics produced by crown bodies, those acting on behalf of crown bodies, or those bodies specified in statutory orders, as defined in section 6 of the Statistics and Registration Service Act 2007.
Public interest/benefit/good	In the context of data and statistics, public benefit includes informing the public about social and economic matters; assisting in the development and evaluation of public policy; and regulating quality and publicly challenging the misuse of statistics. Data sharing and linkage may deliver public benefit at different levels: it may benefit individuals directly, benefit public service providers, and/or bring wider social benefits for people and communities.
Quality	Data and methods that produce assured statistics. Quality means that statistics fit their intended uses, are based on appropriate data and methods, and are not materially misleading. Quality requires skilled professional judgement about collecting, preparing, analysing and publishing statistics and data in ways that meet the needs of people who want to use the statistics.
Safe setting	A service that provides the facilities and support for researchers to securely access sensitive/confidential data. This service may be physical (i.e. located in a specific location that researchers travel to) or virtual (i.e. via remote access from a designated external terminal, or in some circumstances, any terminal that meets minimum security requirements).
Secure Anonymised Information Linkage (SAIL)	The Secure Anonymised Information Linkage (SAIL) platform is a databank of linkable, anonymised datasets about the population of Wales. It contains population, health and social care data.
Social licence	Building public acceptance and approval for something – in this case data sharing and linkage - beyond relying on legal powers.
Synthetic data	Datasets that have the properties of “real” datasets but contain artificial individual-level cases. Their use reduces, but does not necessarily eliminate, the risk of disclosure.

Trustworthiness	Confidence in the people and organisations that produce statistics and data. Trustworthiness is a product of the people, systems and processes within organisations that enable and support the production of statistics and data. Trustworthiness comes from the organisation that produces statistics and data being well led, well managed and open, and the people who work there being impartial and skilled in what they do.
UK Research and Innovation (UKRI)	UK Research and Innovation (UKRI) is the national funding agency investing in science and research in the UK. It brings together the seven Research Councils, Innovate UK and a new organisation, Research England.
Value	Statistics that support society's needs for information. Value means that the statistics and data are useful, easy to access, remain relevant, and support understanding of important issues. Value includes improving existing statistics and creating new ones through discussion and collaboration with stakeholders, and being responsible and efficient in the collection, sharing and use of statistical information.

Annex 2: Organisations that contributed to this review

Over the course of the review, we gathered information, and received feedback on our analysis and recommendations, from people in the organisations listed below. The ideas in the report also benefited from discussions at the UK Data Forum, Government Statistical Service Methodology Symposium, and an OSR event with Directors of Analysis in UK government departments.

- Administrative Data Research Centres: England, Wales, Scotland, Northern Ireland (including the Universities of: Edinburgh; Oxford; Southampton; Swansea; and University College London)
- Administrative Data Research Network
- Australian Bureau of Statistics
- Cabinet Office
- Child and Family Court Advisory and Support Service
- City University
- Cohort and Longitudinal Studies Enhancement Resources (CLOSER), University College London
- Department for Digital, Culture, Media and Sport
- Department for Education
- Department of Health and Social Care
- Department for Work and Pensions
- Education Policy Institute
- Greater Manchester Combined Authority
- Her Majesty's Revenue and Customs
- Institute for Fiscal Studies
- Ministry of Housing, Communities and Local Government
- Ministry of Justice
- National Records of Scotland
- National Services Scotland Information Services Division
- NHS Digital
- Northern Ireland Statistics and Research Agency
- Nuffield Foundation
- Office for National Statistics
- Open Data Institute
- Resolution Foundation
- Royal Academy of Engineering
- Royal Society
- Royal Statistical Society
- Scottish Government
- Secure Anonymised Information Linkage (SAIL) Databank, Wales
- Statistics New Zealand
- The Data Lab
- Wellcome Trust
- Welsh Government
- University of Glasgow
- University of Warwick
- UK Data Service
- UK Research and Innovation – Economic and Social Research Council
- UK Statistics Authority

Annex 3: Materials referenced and other useful sources

Context for the review and general guidance

Administrative Data Research Network: Introduction to Data Linkage (2016)

Administrative Data Taskforce: Improving Access for Research and Policy (2012)

Digital Economy Act 2017

- Provisions for statistical production
- Provisions for research

Nurse Review of Research Councils: Recommendations (2015)

Office for Statistics Regulation: Code of Practice for Statistics (2017)

Public Health Research Data Forum: Enabling Data Linkage to Maximise the Value of Public Health Research Data (2015)

UK Statistics Authority: Better Statistics, Better Decisions: Three Years On (2017)

Public engagement and social licence

Administrative Data Research Network: Developing a Public Engagement Strategy (2018)

British Academy and Royal Society: Data Governance: Public Engagement Review (2017)

Carter, Laurie and Dixon-Woods: The social licence for research: why care.data ran into trouble (2015)

Involve: What Data Should Government Bodies be Allowed to Share About Us? (2014-2016)

Involve: Testing When Data Sharing is for the Public Benefit (2018)

New Zealand Data Futures Partnership: A Path to Social Licence: Guidelines for Trusted Data Use (2017)

NHS Digital: National Data Opt-Out Programme (2018)

Wellcome Trust: Understanding Patient Data programme (2018)

- Guidance on Language
- Spectrum of identifiability
- Case studies of data use

Data ethics and governance

Administrative Data Research Network: Data Reuse for Research Purposes (2017)

British Academy and Royal Society Data Management and Use: Governance in the 21st Century (2017)

Hall and Presenti: Growing the Artificial Intelligence Industry in the UK (2017)

Health and Social care Information Centre: Data Release Review (the Partridge Review) (2014)

House of Lords Artificial Intelligence Committee: AI in the UK: Ready, Willing and Able? (2017)

Information Commissioner's Office: Data Protection Impact Assessments (2018)

Office for National Statistics: The 'Five Safes' - Data Privacy at ONS (2017)

Open Data Institute: What is a Data Trust? (2018)

Reform: Sharing the Benefits – How to use Data Effectively in the Public Sector (2018)

UK Statistics Authority: National Statistician's Data Ethics Advisory Committee (2018)

Data quality, documentation and assurance

Eurostat: ESS Reference Metadata Reporting Standards

FAIR Guiding Principles for Scientific Data Management and Stewardship (2016)

Gilbert et al: GUILD: Guidance for Information about Linking Datasets (2017)

Hand: Statistical Challenges of Administrative and Transaction Data (2018)

UK Data Service: Prepare and Manage Data

UN Economic Commission for Europe: Standards and Metadata

UK Statistics Authority: Quality Assurance of Administrative Data