



Office for  
Statistics Regulation

**Regulatory Guidance**

# **Building confidence in the handling and use of data**

October 2018

# Building confidence in the handling and use of data

Our refreshed [Code of Practice](#) takes a more ambitious approach to considering the range of issues and practices around personal data compared with the first edition of the Code. Rather than solely focusing on the actions for protecting confidentiality, we are promoting an organisational commitment to fulfil the potential of available data.

- 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.

We are challenging organisations to take advantage of the growing opportunity for data integration, stemming from advances in technology and legislation. Sharing and linking data from different administrative sources, or combining with surveys, provide new means for gaining insight about the UK population and to help answer important questions for society. But they also underscore the need for effective and accountable data management.

The wider availability of data about individuals presents new risks in disclosing private information. Consequently, the Code is emphatic in its expectations about ensuring ethical and accountable use, appropriate handling of data, and maintaining security always.

Data are a public good, but the Code also champions the rights of data subjects – we are urging producers to proactively consider the rights of people providing their own data to serve the public good. Respecting these rights empowers citizens in this exchange.

## About this guide

This guide provides some further insight for statistics producers into the principle **T6 Data Governance** when handling data to produce and publish official statistics, and when making that data available to external users, in ways that are transparent and accountable (see page 3).

It also draws on our findings from a review of the UK statistics system's ability to provide greater insight to users via linked data: [Joining Up Data for Better Statistics](#). We have summarised lessons from this review (on pages 4 and 5), to give examples of the actions that producers can take to maximise the potential of data when producing official statistics. They illustrate practices that we look for when assessing compliance with the Code of Practice for Statistics.

We encourage producers to use the **Five Safes** as a framework to inform their practices and communication about keeping data safe, shown on the back page of this guide.

## Spotlight on T6: Data Governance

Data governance sits within the Code's Trustworthiness pillar. **T6 Data Governance** encompasses a wide range of practices involved in data management to support the growing use of data sharing and linkage, while also requiring identifiable data be protected. The need for transparency is emphasised throughout to ensure accountability.

**T6.1** sets the foundation for statistical practice in data governance by emphasising the importance of producer organisations transparently committing to achieving the legal requirements that underpin data collection, handling and release, such as, under:

- [Data Protection Act 2018](#)
- [General Data Protection Regulation 2018](#)
- [Freedom of Information Act 2000](#) and [Freedom of Information \(Scotland\) Act 2002](#)
- [Statistics and Registration Services Act 2007](#)
- [Statistics of Trade Act 1947](#), [Statistics of Trade and Employment \(Northern Ireland\) Order 1988](#)
- [Digital Economy Act 2017](#)

The relevant legal requirements and data ethics standards must be met. National and international guidelines also provide useful approaches for statistics producers – a range of relevant material is given in the [helpful resources](#) section for this principle. Data ethics guidance is available from the Department for Digital, Culture, Media and Sport's [Data Ethics Framework](#), and from the [National Statistician's Data Ethics Advisory Committee](#). These provide practical guidance and a framework for self-assessment. The Open Data Institute has produced a [Data Ethics Canvas](#) which helps organisations to identify and manage data ethics considerations.

**T6.2** requires producers to think about the 'rights' of the living people whose data are being collected. This shifts the emphasis from focusing on data to focusing on the individuals providing their information. It is important for producers to be transparent about their actions and decisions, including about the purposes for collecting the information and how the personal data will be protected. There are some statistical exemptions that can apply when using the data for statistical purposes which have been detailed in [GSS GDPR guidance](#).

**T6.3** focuses on keeping personal data safe and secure – this requires the organisation to ensure a safe setting for data handling at all stages of the statistical production process. These actions require adopting security standards in the technical systems, as well as for staff. It can be best achieved through a culture that promotes a strong awareness of the need for, and ways of, maintaining a safe setting. This practice is closely related to practice T5.4 and the need for providing training on secure data handling – it should be sufficient for the nature of the data use.

**T6.4** addresses the need for the safe release of data and statistics, ensuring that the personal data are appropriately protected, for example by using statistical disclosure control techniques. This can involve making sure that identifiable data are safe, i.e. that they are made anonymous and that published aggregate data cannot be brought together by users to re-identify a person. The practice also requires that safe access is provided to individual-level (micro) data through using protocols such as adopted by the Office for National Statistics' [Secure Research Service](#) and HMRC's [Datalab](#). The practice stresses the importance of transparent processes. We encourage producers to publish materials related to data shares, for example, mandatory and voluntary Data Protection Impact Assessments, public benefit assessments, and records of decisions taken about data share requests. The Wellcome Trust's [guidance](#) in its [Understanding Patient Data](#) programme can help producers explain complex issues, such as, data anonymity and identification risk.

**T6.5** emphasises that solutions should be sought across the organisation, rather than just for individual data sources, and that the solutions should be reviewed regularly. Data management may be subject to frequent change in an evolving environment. It requires a conscious effort to ensure that the arrangements remain robust.

## Achieving a safe and effective data linkage system

Statistics producers should:

- Seek public input to major decisions about statistics & statistical research
- Use clear, consistent & meaningful language
- Be advocates for safe data use
- Produce keeping data safe statements using the Five Safes Framework

Government demonstrates its **trustworthiness** to share and link data through robust data safeguarding and clear public communication

- Maximise opportunities to identify the questions that society wants answered
- Ensure policy maker and external input to question setting process
- Enable more exploratory analysis, including the use of synthetic data

Data sharing and linkage help to answer society's important questions

- Harmonise information governance frameworks
- Increase seniority of sign-off
- Develop risk assessment tools
- Be open and transparent about data shares

Data sharing decisions are ethical, timely, proportionate and transparent

- Design & monitor application processes with users
- Signpost users to other data sources of potential interest

Project proposal assessments are robust, efficient and transparent

- Improve data documentation
- Contribute to an admin data registry
- Use a common framework for documenting and assessing linkage process
- Ensure that statisticians are involved when new data systems are designed
- Ensure data users can provide feedback about data quality to data collectors, to help improve data at source

Data are documented adequately, quality assessed and continuously improved

- Recognise and address resource needs
- Work collectively to address infrastructure requirements
- Adopt creative solutions to reduce data outsourcing costs
- Use effective mechanisms to bring in external expertise
- Support the professional development of data access support staff
- Innovate and share practice around safe settings (especially virtual platforms)

Analysts have the skills and resources needed to carry out high-quality data linkage and analysis

## The Five Safes

### Transparency

To demonstrate trustworthiness, statistics producers must do more than just ensure that robust safeguarding provisions are in place and are followed. There also must be transparency around these safeguarding procedures so they can be scrutinised. To meet this requirement, we expect statistics producers to provide consistent, coherent and accessible information about data safeguarding.

The Five Safes framework developed by ONS is an accessible tool that can be used to help make and communicate decisions about data safeguarding in a consistent and coherent way. It covers five areas.

### 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?

### Proportional to the situation

The steps that will need to be taken to address each of the five areas will vary depending on the type of data and how they are used. For example, 'safe people' will include the IT security training provided by organisations to all staff, and the minimum qualifications expected of government analysts, as well as the protocols to approve external researchers for accessing microdata in secure settings, if applicable. 'Safe settings' covers many aspects, including buildings and IT system security provided by an organisation for its staff, and the additional security procedures that apply in microdata labs for external users, either in person or via remote access. Organisations will need to use their own judgement to decide what steps need to be taken in each area. The framework provides a way of ensuring that all necessary aspects have been considered.

### Prominence

We want statements about keeping data safe to be as prominent and commonplace as statements about data quality. Using the Five Safes as the template for these statements will help provide some consistency and coherence across statistics producers.