

## Data Linkage work programme

This work programme has been produced in response to the Office for Statistical Regulation (OSR) report 'Joining Up Data for Better Statistics', which can be found here: <https://www.statisticsauthority.gov.uk/publication/joining-up-data/>.

The work programme is set out under the six themes of the OSR report. It is intended to be read alongside that document and is further complimented by the National Statistician's response to the report, which is published with this document. The work programme describes the system-wide enabling work that departments are undertaking to facilitate better linking of data across government, in addition to the good practice being undertaken within individual departments that is highlighted in Annex B.

### **1: Government demonstrates trustworthiness, robust safeguarding and clear public communication.**

The legal and ethical safeguards which ensure its use for public interest underpin the development of an efficient and dependable data linkage system. Having the correct data architecture including data management and governance frameworks in place, is an integral foundation for maximising the use of data linkage, whilst ensuring that peoples' information is kept safely and securely.

The Office for National Statistics (ONS) has initiated and lead the cross-government Data Architecture Network which works across the public sector to provide strategic direction to standardise the way we work and ease communication between government departments. More details are available at: <https://dataarchitecture.blog.gov.uk/>. Alongside this activity ONS Data Architecture are developing a Data Strategy that will enable us to put in place the correct data capability infrastructure, governance and management framework to serve the statistics and research community in the future, which both extracts value from data and has the appropriate safeguards which assure public confidence. The data management framework that underpins the Data Strategy will comprise:

- i) Data Principles – a set of high-level principles that define the scope and path of the data journey, and form the guiding framework for data management. The data principles can be viewed at varying levels of detail to meet the needs of a range of users, from the lay person to technical specialists. They will be published by early 2019.
- ii) Data Policies - statements of intent that support the Data Principles, implemented as procedures or protocols. ONS is now reviewing all of its data and statistical policies, and has developed a range of new policies to reflect changes in the data landscape. For example, ONS recently published a web scraping policy. ONS will be publishing a holistic view of our data and statistical policies in early 2019. This will continue to be reviewed to ensure our policies remain up to date and fit for purpose.
- iii) Data Standards - a set of mandatory rules which set out how systems are configured and operated in support of the Data Policies. The standards provide detailed instructions on what ONS data architecture looks like and align to existing best practice such as international standards. ONS has developed a range of new data standards to define our new data architecture. Additional, standards will continue to be developed and reviewed to ensure they evolve as technology and circumstances change.

Action:

- ONS will publish the key components of the Data Strategy and management framework (including a list of data holdings) in early 2019. This will be promoted across the Data Architecture network and the statistics and research community.
- ONS will engage across the statistics and research community on the high-level principles within the Data Strategy to understand detailed departmental requirements for policy development.

## **2: Data sharing and linking help to answer society's important questions.**

Statistics must keep pace with a fast-changing world to answer society's important questions. Collaboration and data linkage are critical to the success of the statistical system in responding rapidly to emerging needs.

The Data Advisory Board, which includes senior data leaders across government, has agreed a strategic focus on four areas of data access, quality, use and capability and has tasked departments to develop a work programme that will inform the National Data Strategy. ONS is actively contributing to this work, for example in developing a set of high-level data infrastructure principles as part of the ONS Data Strategy outlined in section 1. Departmental Heads of Profession for Statistics are also active members of other cross-government networks.

The Government Statistical Service (GSS) data project has already begun work on common infrastructure to increase the organisation and integration of published GSS statistical data. The project will develop dataset families to link groups of related, aggregated data into a collection, so that users can better discover and navigate our statistical data, helping them answer society's important questions. This facility has already been delivered for Trade statistics in the initial phase of the project. More information can be found here: <https://gss.civilservice.gov.uk/guidance/the-gss-data-project/>.

Action: Incrementally increase and publish cross-GSS dataset families available each year.

Under this theme, the Office for Statistics Regulation have highlighted the need for more exploratory analysis to take place across the GSS. I support this. ONS is recommending greater use of research using synthetic data across the GSS via the National Statistician's Quality Review (NSQR) on Privacy and confidentiality (<https://gss.civilservice.gov.uk/guidances/quality/nsqr/privacy-and-data-confidentiality-methods-a-national-statisticians-quality-review/>) and is already leading a Synthetic Data Pilot project of which Phase 1 was recently completed. A summary report will be published in January 2019, while Phase 2 began in summer 2018.

Action:

- ONS to publish report on Phase 1 of the Synthetic Data Pilot project in January 2019.
- Establish a GSS task and finish group by March 2019 to implement and publish the recommendations from the NSQR.

## **3: Data sharing decisions are ethical, timely, proportionate and transparent.**

To make the most of the opportunity presented by the data revolution, we must respect and protect privacy and confidentiality.

The National Statistician's Data Ethics Advisory Committee (<https://www.statisticsauthority.gov.uk/about-the-authority/committees/nsdec/>) addresses this need. The committee is made up of experts from academia government, and civil society, to independently advise the National Statistician on the ethical dimensions associated with the access, use and sharing of public data, for research and statistical purposes. The committee considers projects from across the GSS and wider government, academia and the commercial sector.

To enhance the ethical evaluation of proposals to link and share data for statistics and research, the OSR review recommends that independent and appropriately skilled people conduct ethical assessments. The ethical framework developed by the National Statistician's Data Ethics Advisory Committee (<https://www.statisticsauthority.gov.uk/about-the-authority/committees/nsdec/data-ethics/>) is an important tool here. This can be further facilitated by senior involvement and the relevant senior officer in departments responsible for signing data sharing agreements should satisfy themselves that such appropriate ethical assessments have been undertaken, and any ethical risks have been mitigated.

**Action:** ONS to recommend that data-sharing agreements approved at senior levels across government take sufficient account of ethical considerations and frameworks as part of the review and approval process.

In operating the statutory framework for statistics and research set out in the Statistics and Registration Service Act 2007, the Digital Economy Act 2017, the Data Protection Act 2018, and other relevant legislation it is important to minimise the burden on data providers of linked data by maximising collaboration and engagement, and considering any resource and related implications. ONS seeks to do this by assisting departments through secondments, feasibility subsets and other modes of assistance. The Data Advisory Board may also play a role here in considering options around identifying additional financial and resource opportunities which may be available.

The UK Statistics Authority (UKSA) has already published guidance for Heads of Profession for Statistics, and the wider GSS, in the application of data protection law for statistical work brought about by major changes after the introduction of the General Data Protection Regulation (GDPR) <https://gss.civilservice.gov.uk/policy-store/uk-statistics-authority-guidance-on-the-general-data-protection-regulation-gdpr/>.

#### **4: Project proposal assessments are robust, efficient and transparent.**

The OSR review highlights the importance of advocating safe data use and that projects pursued must serve public benefit across government and within the statistics and research community.

ONS is already upgrading the Secure Research Service (SRS) to provide, researchers with a private cloud-based portal accessible from their desktops, subject to security protocols. At the same time the Digital Economy Act makes it easier for departments to share data and will create new opportunities to link data. This is likely to result in more requests from researchers for access to linked data.

To ensure security and transparency ONS has developed the Five Safes Framework in which accredited statisticians and researchers must work in line with 'keeping data safe' statements (<https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/>.) ONS also provides training for

all researchers wishing to use unpublished, record-level, Government data through the DEA (and other legal gateways), on how to ensure safe use of data through the Five Safes framework. ONS publishes details of all the datasets that are shared with external researchers.

The UKSA independently accredits all researchers, projects, and organisations who process or deidentified data and are provided secure access environments for research purposes. All organisations and individuals must demonstrably meet the standards for accreditation set out in the statutory research code of practice and accreditation criteria, including standards on data ethics, professional skills and capability, data security, proportionality and transparency, to maintain public trust and confidence in how data is being used to support UK research for the public good.

Action:

- End to end testing of the upgraded SRS is planned for early 2019, and following this access to the new service will be rolled out in the Spring.
- ONS will engage bilaterally with Departments, early in 2019, to discuss the DEA, and how the SRS can be used to enable secure research access to data from across Government, without the need for others to develop parallel services.
- ONS advocate the Five Safes Framework and are happy to provide advice on its use across the GSS and more widely.

## **5: Data are documented adequately, quality assessed and continuously improved.**

Data catalogues document and keep track of all the data sources used by an organisation. A catalogue should have an entry for each data source and maintain details about how they are used, including details such as who has access, how they are combined, what attributes they contain and the responsible owner.

ONS is committed to and leading on developing and implementing appropriate data management practices. For data sources inside ONS, this information is currently held in the Information Asset Register. The Data Strategy described in Theme 1 will be underpinned by a new version of the Catalogue and Metadata standards. ONS recognises the urgent requirement to determine what needs to be stored in the catalogue and to deliver this is a joined-up way to address future requirements. Other Government departments have similar solutions relating to their ICT strategies.

Action:

- ONS Discovery into Catalogue and Metadata standards begins January 2019.
- Planned development of a prototype of the Catalogue and Metadata standards in 2019/20.
- Best practice on use of catalogues will be shared across the Data Architecture network

ONS will work in collaboration with government departments, researchers, analysts and stakeholders to access and document data in a streamlined and efficient way. To achieve this ONS has begun a research programme to develop administrative data methods. This will drive collaboration in the development of new statistical frameworks and systems, including developing methods to link multiple sources for several uses, linkage of big data and unstructured data, and the development of linkage quality measures.

Action:

- Details about the first projects in the research programme will be published in January 2019.

- ONS will seek further input on priority research questions across the GSS in January 2019.

Currently the GSS Methodology Advisory Service (MAS) <https://gss.civilservice.gov.uk/wp-content/uploads/2018/04/MAS-brochure-v1-FINAL.pdf> and GSS Methodology Advisory Committee (MAC) <https://gss.civilservice.gov.uk/about-us/governance/methodology-advisory-committee/> provide free methodological advice to support quality improvements and enable access to subject matter experts inside and outside government, including from world leading experts in data linkage methods. Recent changes to the GSS MAC have enabled better access to cutting-edge methods and made it easier to engage with other experts across different sectors, maximising benefits for the wider analytical community.

ONS leads on the NSQRs <https://gss.civilservice.gov.uk/guidances/quality/nsqr/privacy-and-data-confidentiality-methods-a-national-statisticians-quality-review/> which cover thematic topics of national importance, conducted on behalf of and for the GSS. These reviews complement existing quality assurance practices, offering an additional tool to make sure methods remain fit for purpose and among the best in the world. They provide an opportunity for experts outside the GSS to contribute to the continued improvement of data linking methods and support the GSS identify what good practice looks like for these methodologies as well as help identify opportunities for further development and investment.

## **6: Analysts have the skills and resources needed to carry out high quality data linkage and analysis.**

The right skills and resources available in the statistical system will be essential to deliver the outcomes we need.

We are seeking to strengthen capability through the Analysis Function Strategy (2018) (<https://www.gov.uk/government/publications/the-analysis-function-strategy>), which will broaden analytical capability across government including for the GSS. Harnessing the opportunities from data linking and data science requires collective expertise from across disciplines which is why Statisticians are collaborating with Operational Researchers, Social Researchers, Economists, Scientists, Engineers, Actuaries and Digital Technology specialists to build a coordinated Learning and Development offer for analysts in government.

Alongside this there will be an updated and expanded professional development programme which draws on expertise from outside of government. This will promote an effective and multiskilled workforce to respond to emerging needs and assure user confidence.

Action:

- Analysis Function Standard to be launched in 2019.
- The statistics profession will pilot e-learning across the GSS in 2019

The UK statistical system will continue to build its analytical capability, using external collaboration opportunities and sharing internal good practice, to enhance staff understanding, confidence and capability with regards to safe data linking and its approach to consistent data standards.

To achieve this, we have invested in ONS to engage and collaborate to address the methodological and capability needs of users and producers when linking data, whilst also identifying examples of good

practice to ensure they are being replicated across the rest of the GSS. ONS will provide training to develop and support a skills strategy across the GSS, and create a community of best practice.

Action – ONS will make data linkage methods training material and guidance available and across the GSS by March 2019. ONS will explore e-learning tools to increase reach across government.

There is already a wealth of existing work across government to develop robust and trustworthy data linkage. To strengthen co-ordination, the Best Practice & Impact division of ONS will organise a data linking conference in 2019. This will bring together all the key players from government and external organisations, to further progress recommendations from the OSR report, understand departmental requirements, share best practise, strengthen relationships between organisations and increase statistical capability.

Action – ONS to organise a government wide data linking conference in 2019.