CLOSER's UK National Data Strategy consultation response

CLOSER, the home of longitudinal research

CLOSER, the home of longitudinal research, brings together world-leading longitudinal population studies to maximise their use, value and impact and improve our understanding of key social and biomedical challenges. CLOSER does this by:

- Enabling change over time and across generations to be better understood.
- Enhancing insights through data linkage.
- Helping researchers find data.
- Equipping the next generation of scientists.
- Supporting innovative research projects.
- Maximising the impact of longitudinal research.

The following studies are part of CLOSER:

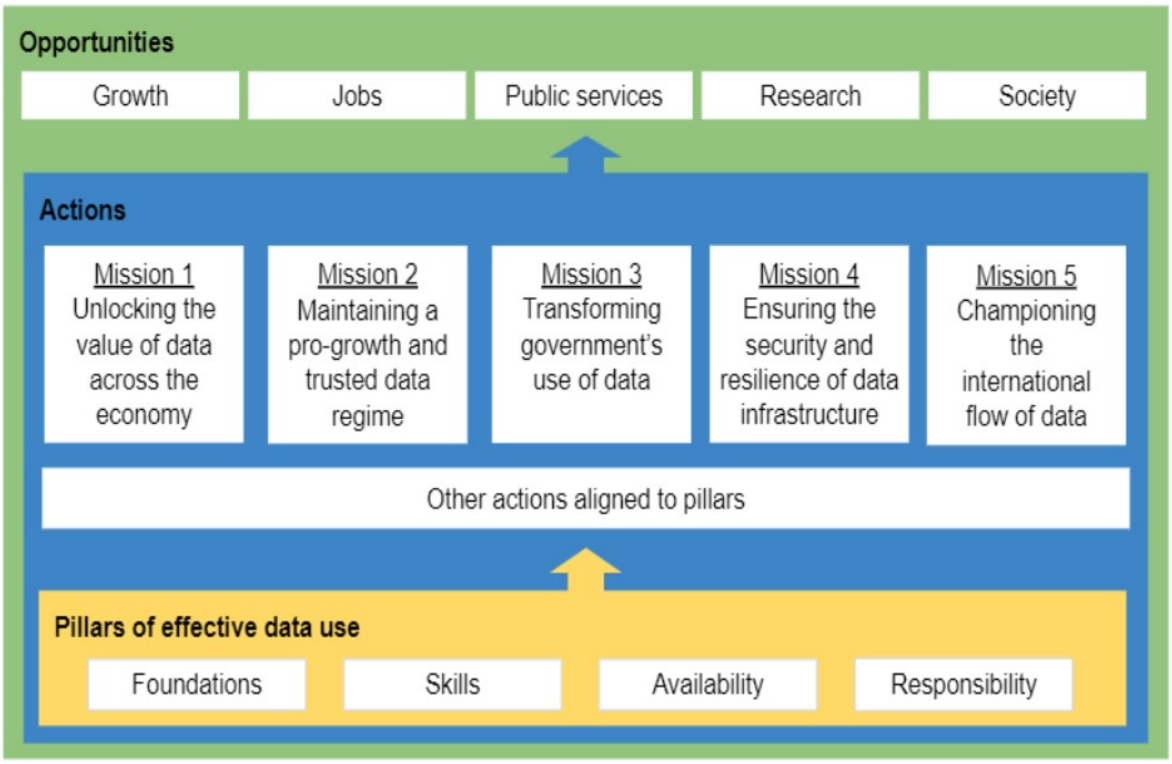
- Hertfordshire Cohort Study (University of Southampton)
- MRC National Survey of Health and Development (UCL)
- 1958 National Child Development Study (UCL)
- 1970 British Cohort Study (UCL)
- ONS Longitudinal Study (Office for National Statistics)
- Whitehall II Study (UCL)
- Understanding Society: The UK Household Longitudinal Study (University of Essex)
- Avon Longitudinal Study of Parents and Children / ALSPAC (University of Bristol)
- Southampton Women's Survey (University of Southampton)
- Millennium Cohort Study (UCL)
- English Longitudinal Study of Ageing (UCL)
- Next Steps (UCL)
- Growing Up in Scotland (The Scottish Government)
- Generation Scotland (University of Edinburgh)
- Born in Bradford (Bradford Teaching Hospitals NHS Trust)
- Wirral Child Health & Development Study (University of Liverpool)
- NICOLA (Queen's University Belfast)
- LSYPE 2 (Department for Education)
- Health and Employment After 50 / HEAF (University of Southampton)

CLOSER is funded by the UKRI Economic and Social Research Council (ESRC) and is based at the UCL Social Research Institute: www.closer.ac.uk

Overall

We want to ensure that we produce a forward-looking strategy that takes into account public opinion and delivers real change. These questions will help to inform future work that the government will take in this space. They will provide evidence for the government to target areas for intervention in future policy.

Please find a diagram below of the NDS pillars, missions and opportunities for reference.



Q1. To what extent do you agree with the following statement: Taken as a whole, the missions and pillars of the National Data Strategy focus on the right priorities. Please explain your answer here, including any areas you think the government should explore in further depth.

Strongly disagree
Somewhat disagree
Neither agree nor disagree
Somewhat agree
Strongly agree

Q1. CLOSER response:

Data Foundations
The National Data Strategy identifies a number of areas and issues that CLOSER has previously highlighted that require addressing: these include issues around data

quality, the need for common standards, and use of metadata. We would like to particularly emphasise that the provision and consistent use of metadata is fundamental for efficient and meaningful exchange of data between organisations.

Standards

To maximise the use of data across the UK, there is a need for comparable ontologies, for example European Social Science uses the same terminology with translations so that you can locate the same information across different European archives – this approach needs to be adopted in the UK. The adoption of the recent suite of metadata standards (see: https://www.ons.gov.uk/aboutus/transparencyandgovernance/datastrategy/datastandards) could be transformational in not only providing the basis for efficient exchange of data, but also to combine data meaningfully.

The recommendation by the Office for National Statistics (ONS) to use the Asset Description Metadata Schema (ADMS) would (if adopted widely) be a major step forward in allowing data to move seamlessly within government.

Agreement on the structure and format is not in of itself sufficient to achieve such an ambition. The European Commission's adoption of the ADMS recognised the need for a co-ordinating function for rationalising the terminologies and put in place a Metadata Registry to co-ordinate these terms and vocabularies. This has been instrumental in ensuring data quality and the reliability of data collection. The current Registers (https://www.registers.service.gov.uk) is a useful start, but would need to rapidly expand (beyond its current 50 items) if the current ambitions are to be realised.

Metadata

The area of metadata is fast moving and would need to intersect with the changes in technology occurring in government more widely, for example, the adoption of cloud and big data technologies

We see the need for a 'Metadata Strategy Office' or similar such body that has the power, authority and expertise to assess and recommend standards, co-ordinate the activities of a Metadata Registry and champion good practice through outreach and training activities.

Data Skills

We agree that data skills are a fundamental pillar of effective data use and should remain a priority of the National Data Strategy. This is an area we feel requires more focus: CLOSER's training needs review (https://www.closer.ac.uk/wp-content/uploads/CLOSER-Analytical-training-needs-review-full-report.pdf) identified several fundamental issues in the level of data skills across both research and policy domains. These include: a lack of specific analytical skills, inability to access or handle data, a lack of training provision for specific techniques, software literacy, physical access to data (for example, those held in a secure space), and poor awareness of the complexity of longitudinal data. Furthermore, reviews by both the Economic and Social Research Council (ESRC) and CLOSER highlighted gaps in understanding, accessing and using longitudinal data, particularly by government users. Increasing training provision and removing barriers to training access should

be explored in further depth by the government to ensure that users are better equipped to analyse, interpret and understand a wide range of data, including longitudinal data.

We welcome the commitment to publish a working definition of data skills for the wider economy, however would emphasise the need for this to recognise the very different types of data that exist and range of (in some cases) very specific and technical skills are required.

Case study: CLOSER Discovery
CLOSER Discovery - the UK's most detailed search engine for longitudinal data - enables researchers and analysts to search through the rich metadata from UK longitudinal population studies and see, at a glance, which studies meet their research requirements.

It is a user-friendly resource for locating the variables that best suit research interests, and testing their robustness. The platform is built using metadata. By using metadata, CLOSER Discovery enables users to find out what is available in the UK's longitudinal datasets, understand the context of how the data were created, and make an assessment about whether the data is relevant for research.

CLOSER Discovery demonstrates how you can utilise and extend existing standards in both data and metadata to build new capacity, products and services that fill an unmet need, while also enhancing existing skills and capacity within the longitudinal studies and research community.

Access CLOSER Discovery: https://discovery.closer.ac.uk

Q4. We welcome any comments about the potential impact of the proposals outlined in this consultation on the UK across all areas, and any steps the government should take to ensure that they take account of regional inequalities and support the whole of the UK?

Q4. CLOSER response:

CLOSER has recently expanded to include more longitudinal population studies, bringing together 19 national and regional studies. Part of the rationale for this is in recognition of the importance of 'place' and the need to better understand regional differences (this is particularly relevant in light of the COVID-19 pandemic). More regional longitudinal population studies, such as Born in Bradford, the Avon Longitudinal Study of Parents and Children, and Wirral Child Health and Development Study, should be created in key regions to help understand regional inequalities across the UK and generate new insights. It is worth noting that new studies have recently been set up in Liverpool and East London: Children Growing Up in Liverpool (C-GULL), which is tracing the lives of over 10,000 Liverpudlians to understand more about what influences the health and wellbeing of children and their families living in the region and East London Genes and Health, a long-term study of 100,000 people of Bangladeshi and Pakistani heritage which aims to improve health among these people and help fight heart disease, diabetes, and poor

health in the UK. The data from these new studies will generate valuable insights into specific regions and groups of the population in the UK, providing robust scientific evidence to help the government understand and take steps to reduce inequalities.

Mission one: Unlocking the value of data across the economy

Data is an incredibly valuable resource for businesses and other organisations, helping them to deliver better services and operations for their users and beneficiaries. However, there is increasing evidence to suggest that the full value of data is not being realised because vital information is not getting to where it needs to be.

Our first mission is to create an environment where data is appropriately usable, accessible and available across the economy – fuelling growth in organisations large and small. We will create a clearer policy framework to identify where greater data access and availability across and with the economy can and should support growth and innovation, in what form, and what government's role should be, in the UK and globally.

Data availability: For data to have the most effective impact, it needs to be appropriately accessible, mobile and re-usable. That means encouraging better coordination, access to and sharing of data of appropriate quality between organisations in the public sector, private sector and third sector, and ensuring appropriate protections for the flow of data internationally.

Q5. Which sectors have the most to gain from better data availability? Please select all relevant options listed below, which are drawn from the Standardised Industry Classification (SIC) codes.

Accommodation and Food Service Activities
Administrative and Support Service Activities
Agriculture, Forestry and Fishing
Arts, Entertainment and Recreation
Central/Local Government inc. Defence
Charity or Non Profit
Construction
Education
Electricity, Gas, Steam and Air Conditioning Supply
Financial and Insurance Activities
Human Health and Social Work Activities
Information and Communication
Manufacturing
Mining and Quarrying
Transportation and Storage
Water Supply; Sewerage, Waste Management and Remediation Activities
Wholesale and Retail Trade; Repair Of Motor Vehicles and Motorcycles
Professional, Scientific and Technical Activities
Real Estate Activities

Other

Q6. What role do you think central government should have in enabling better availability of data across the wider economy?

Q6. CLOSER response:

At a minimum, central government should have a role in setting the parameters that means that data is available to citizens in a usable and comprehensible form.

A good example of where that has been achieved is the Data Communications Company (DCC) established by OFGEM to ensure transparency of data from smart meters for both energy providers and consumers.

Early results from the smart meter data coming out of the DCC indicate that the standards may have been too relaxed (resulting in 65% of households having more than 50% usable electricity readings) [http://doc.ukdataservice.ac.uk/doc/8643/mrdoc/pdf/8643_serl_smart_meter_data_quality_report_v2020_07.pdf].

The work of the DCC illustrated an important area in setting standards that also ensures meaningful data portability. Exporting the data as plain text is in the strictest terms data portability, but unless it can be used in another system it has little real utility.

Central government should also have a role in ensuring that there are robust reviews of the processes by which data and metadata standards are set and implemented in these complex projects. It is important that the performance of such standards are assessed at an early stage against their original criteria to help inform and improve similar activities in the future.

Data foundations: The true value of data can only be fully realised when it is fit for purpose, recorded in standardised formats on modern, future-proof systems and held in a condition that means it is findable, accessible, interoperable and reusable. By improving the quality of the data we are using, we can use it more effectively, and drive better insights and outcomes from its use.

Q7. To what extent do you agree with the following statement: The government has a role in supporting data foundations in the wider economy. Please explain your answer. If applicable, please indicate what you think the government's enhanced role should be.

Strongly disagree
Somewhat disagree
Neither agree nor disagree
Somewhat agree
Strongly agree

Q7. CLOSER response:

Considerable value emerges when we link the rich and extensive information collected about individuals through their participation in longitudinal population studies with the detail that comes from administrative data held about the same people. However, the potential offered by survey and administrative data linkage are often hindered by cultural barriers within government departments and disparate and opaque data access processes and procedures. In our experience, there are a number of barriers to accessing data held by government, including legislative uncertainty, resource (including insufficient staff with the necessary data management skills and high staff turnover) and out of date data management structures or procedures.

There have been a number of recent activities which have brought together practitioners such as the 'Data Bites' series organised by the Institute for Government (IfG), the NHS England and Improvement series of virtual huddles and other informal data huddles to bring together a wide range of actors within the data community and highlight and share good practice. If these initiatives are to be sustained, they need to be embedded as standard ways of working. Data champions situated within government departments may be a way of helping to change and solidify a change in culture, identify the specific barriers in individual departments and work together to address these. These would have to be at an appropriate level and given the authority to effect change – our experience with the Longitudinal Studies Champions within government departments is that these were not at an appropriately senior level, nor were they empowered to champion the use and understanding of longitudinal survey data across their respective departments. Churn within the civil service was also a major factor affecting these roles, with long gaps between the postholder leaving their role and a new champion being appointed.

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

There is massive untapped potential in the way the government uses data. We will implement major and radical changes in the way that the government uses data to drive innovation and productivity across the UK. In doing so, we will improve the delivery of public services, as well as our ability to measure the impact of policies and programmes, and to ensure resources are used effectively.

To succeed, we need a whole-government approach led by a Government Chief Data Officer from the centre in strong partnership with organisations. We need to transform the way data is collected, managed, used and shared across government, including with the wider public sector, and create joined-up and interoperable data infrastructure. We need the right skills and leadership to understand and unlock the potential of data – and we need to do so in a way that both incentivises organisations to do the right thing, as well as build in

the right controls to drive standardisation, consistency and appropriate data use.

The government is going to set an ambitious package of work in this space and wants to understand where we can have the biggest impact.

Q12. We have identified five broad areas of work as part of our mission for enabling better use of data across government:

Quality, availability and access
Standards and assurance
Capability, leadership and culture
Accountability and productivity
Ethics and public trust

We want to hear your views on any actions you think will have the biggest impact for transforming government's use of data.

Q12. CLOSER response:

Quality, availability and access
The adoption of big data technologies has the potential to open up both efficiencies and offer new insights into data across government. It also has the potential to create a vast pool of incomprehensible information.

Good metadata management is increasingly recognised as being a key part of a strategy to truly unlock the potential of such data. Too much time and effort is currently being wasted trying to clean and understand such data. Good data standards and co-ordination of metadata are key to being able to extract real value from such increasingly important data sources.

Public Trust
The relationship with participants in the CLOSER longitudinal population studies span across a number of decades, presenting a unique opportunity for others to understand how people view their data being used for research purposes. The CLOSER studies invest considerable resource to understand public expectations and to maintain participant trust, for example, studies use public and participant feedback to steer the development of data security and safeguarding facilities that allow research on linked routine records while maintaining legal compliance and public acceptability. We would urge learning from the experience from these long-standing relationships and that the measures longitudinal studies population have in place should be considered when developing plans around trustworthiness of data collection, usage and how these are communicated to study participants and the wider public.

Q13. The Data Standards Authority is working with a range of public sector and external organisations to create a pipeline of data standards and standard practices that should be adopted. We welcome your views on standards that should be prioritised, building on the standards which have already been recommended.

Q13. CLOSER response:
As we have previously noted, setting appropriate and well-defined standards is a crucial first step in unlocking the value of data.

The priority should now be on adoption, whilst ensuring quality. NHS Digital's experience of creating centralised Hospital Episode Statistics (HES) datasets from the many NHS providers is worth considering in this regard: obtaining data in a regularised format from all providers was prioritised, and providers were then encouraged and incentivised through quality metrics to align with centrally agreed terminology, to allow the data to be more easily and reliably compared and combined.