

Collections development policy



22 February 2023

Public

Copyright © 2023 University of Essex. Created by UK Data Archive, UK Data Service.

Version: 10.00



Table of contents

1. Scope	3
2. Definition of terms	3
3. Introduction	4
4. Background and context	4
5. Policy principles	5
6. Scope of collections.....	7
6.1 Disciplines.....	7
6.2 Strategic values	8
6.3 Out of scope materials.....	8
7. Selection and appraisal criteria	9
8. Curation categories.....	10
9. Licence and access framework.....	10

1. Scope

The mission of the UK Data Service is to support high quality social and economic research, teaching and learning through assuring long term access to quality economic, population and social research data, supporting and promoting their use, value and impact.

The Policy describes the functions and activities implemented by the UK Data Service for developing its data collections to ensure the successful delivery of its mission. These functions and activities include the acquisition, selection and appraisal of data collections while aligning with and meeting the data needs of our major stakeholder communities, other relevant organisational policies (both internal and external), national and international standards and legislative requirements.

2. Definition of terms

Acquisition: The process related to acquiring materials for inclusion in the collection.

Archiving: A curation activity that ensures that data are properly selected, stored, can be continuously accessed, and that its logical and physical integrity is maintained over time, including authenticity and security.

Data Owner: The natural or legal person, public authority, agency or any other body which alone, or jointly with others, holds the copyright and associated intellectual property rights in a Data Collection.

End User Licence Agreement: The User Agreement entered into by a User when registering to access Safeguarded and Controlled data from the UK Data Service.

Ingest: The processes related to receiving information from an external source and preparing it for archival storage and access.

Personal Data: Are defined as in accordance with the UK General Data Protection Regulation (UK GDPR) Article 4(1) and the Data Protection Act (DPA) 2018 s3(2) as: data that relate to an identified or identifiable natural person, be it directly or indirectly, taking into account other information derived from published sources.

Personal Information: Information that relates to and identifies an individual (including a body corporate) taking into account other information derived from published sources (as defined in section 39(2) of the Statistics and Registration Service Act 2007).

Preservation: An activity within archiving in which specific items of data are maintained over time so that they can still be accessed and understood through changes in technology. Unlike 'archiving', preservation does not include maintaining the logical and physical integrity of a data object.

Selection: The process of identifying materials to be made accessible and in many cases to be preserved for long-term use.

User Agreement: An agreement setting out the terms and conditions of data use and establishing the rights and responsibilities of the Registered User.

3. Introduction

The UK Data Service is a domain specific repository acquiring data for its designated user community which mainly consists of social science and related data users within the Higher and Further Education sector in the UK. However the UK Data Service services are designed to be easily accessible to all types of data user with basic understanding of social science methods and techniques relevant to the data collections being accessed.

The aims of the Policy are:

- To define the Policy drivers and scope of the UK Data Service collection.
- To provide an overview of the criteria for, and principles behind, inclusion of data.
- To steer the selection and appraisal of data collections and to ensure a consistent and transparent approach.
- To inform key stakeholders including existing and prospective data owners and data users, funding bodies and partners of the applicable criteria for data collections that are to be acquired.
- To inform the effective resourcing needed for the implementation of key collections development activities.

This Policy directly informs the key service requirements of the UK Data Service and significantly impacts on the UK Data Archive's Preservation activities.¹

4. Background and context

The primary policy driver is the UK Research and Innovation Economic and Social Research Council (UKRI ESRC) Research Data Policy.² This policy demonstrates the UKRI ESRC's commitment to the provision of access to data as the primary asset for economic and social research regardless of its underlying access criteria (section 7). The principles are guided by the earlier OECD's report Principles and Guidelines for Access to Research Data from Public Funding and the UK Strategy for Data Resources for Social and Economic Research.^{3,4} The policy is also informed by the Office for National Statistics (ONS) research and data access

¹ <https://dam.data-archive.ac.uk/controlled/cd062-preservationpolicy.pdf>

² <https://www.ukri.org/publications/esrc-research-data-policy/>

³ <https://esrc.ukri.org/files/research/uk-strategy-for-data-resources-for-social-and-economic-research>

⁴ <https://www.oecd.org/sti/inno/38500813.pdf>

policy⁵ and the 2017 Digital Economy Act Research Code of Practice and Accreditation Criteria⁶ that make a commitment to the provision of access to ONS microdata for statistical research. This strategy is also enshrined in the Concordat between the ONS and the UK Data Service. The UK Census 2021/22 programmes further offers the opportunity to provide access to enhanced data products under this policy for the research community.^{7,8,9}

There are many external policies and legal frameworks, which may indirectly impinge on the implementation of this Policy. These include legislation, e.g. the Census Act 1920; the Statistics and Registration Service Act 2007, §39(5); The Digital Economy Act 2017; the Data Protection Act 2018; the United Kingdom General Data Protection Regulation (UK GDPR); and the General Data Protection Regulation (EU GDPR), international standards, e.g. ISO 27001; and HMG Infosec documents.

The Policy demonstrates a commitment to aligning the data needs of our major stakeholder communities: data users, data owners and data funders and ensuring that those needs can be met by the data service infrastructure provided by the UK Data Service. The Policy is intended to be flexible enough to meet the challenges and changing demands from these communities in an ever rapidly evolving data landscape.

The Policy shall be responsive in both meeting the demands of changes in technology and resulting new types of data. The UK Data Service will advance the mission of the UKRI ESRC by investigating and supporting the deposit of new types of data and new resources in the collection in response to current and emerging research themes and user needs.

5. Policy principles

The policy is founded on the following principles:

1. Advancing the strategic objectives of the UKRI ESRC.
2. Contributing to the successful delivery of the UKRI ESRC Research Data Policy.
3. Contributing to the successful delivery of the UKRI ESRC Strategic Delivery Plan, 2022-2025 and subsequent post-2025 plans.
4. Supporting the designated user community by providing access to a wider range of data sources and resources, regardless of access method (i.e. Open, Safeguarded,

5

<https://www.ons.gov.uk/aboutus/transparencyandgovernance/datastrategy/datapolicies/onsresearchanddataaccesspolicy>

⁶ <https://www.gov.uk/government/consultations/digital-economy-act-part-5-data-sharing-codes-and-regulations/research-code-of-practice-and-accreditation-criteria>

⁷ <https://www.ons.gov.uk/census>

⁸ <https://www.nisra.gov.uk/statistics/census/2021-census>

⁹ <https://www.scotlandscensus.gov.uk/about/>

Controlled).

5. Ensuring that data collections brought into the service are of the highest quality.
6. Promoting and supporting high-quality basic, strategic and applied research in the social sciences.
7. Promoting teaching and learning use of data and other resources to encourage secondary usage.
8. Strengthening existing partnerships with data owners and producers whilst establishing new collaborative partnerships and creating new opportunities.

Commentary

Principles 1 and 2: The UK Data Service as a data service provider is mandated by the UKRI ESRC through its Research Data Policy to provide advice and guidance to award holders on implementing effective data management plans. UK Data Service will provide guidance to award holders on issues relating to confidentiality, security and ethics in data management and sharing and work closely with other UKRI ESRC investments such as the Cohort and Longitudinal Studies Enhancement Resources (CLOSER) to ensure the discoverability of UKRI ESRC funded data assets.

Principle 3: The vision of the UKRI ESRC Strategic Delivery Plan, 2022-2025¹⁰ is guided by four principles for change: diversity, resilience, connectivity and engagement. This policy will reflect the objectives of the strategy while continuing to build on good scientific training and ‘breadth and depth’ of readily available resources.

Principle 4: This work continues to support the long established principle and commitment of the UKRI ESRC to ensure that publicly funded research data are a public good, produced in the public interest and should be openly available to the maximum extent possible. The UK Data Service will strive to maximise the access to any data it collects while meeting legislative and legal requirements.

Principles 5, 6 and 7: The UK Data Service will ensure the authenticity, reliability and logical integrity of all resources entrusted to our care while providing usable versions for research and teaching or learning in perpetuity. The UK Data Service will ensure that its user community has access to the materials it requires by continuing to maintain existing resources, and the currency of its major data series.

Principle 8: The UK Data Service will work with European partners to provide and facilitate

¹⁰ <https://www.ukri.org/wp-content/uploads/2022/09/ESRC-010922-StrategicDeliveryPlan2022.pdf>

access to international data in both quantitative and qualitative forms. We will work to facilitate new collaborative partnerships and explore opportunities to contribute to the development of federated resourced discovery efforts and methods and tools for data curation and exploration.

6. Scope of collections

The UK Data Service collection is a wide-ranging resource of great value to research and teaching communities across and beyond the social sciences in the UK and internationally.

The collection includes but is not limited to large-scale government and ESRC-commissioned survey and aggregate data, international macrodata, qualitative and mixed methods data, as well as other data created with UKRI funding.

6.1 Disciplines

Supporting the implementation of the UKRI ESRC Research data policy, remit and portfolio¹¹, and recognizing the growing importance of supporting cross-disciplinary research, UK Data Service selects, acquires and disseminates data from a wide range of disciplines including but not limited to:

- area and development studies
- demography
- economic and social history
- economics
- education
- environmental planning
- human geography
- international relations
- linguistics
- management and business studies
- politics
- psychology

¹¹ <https://www.ukri.org/councils/esrc/remit-portfolio-and-priorities/>

- science and technology studies
- social anthropology
- social policy
- social statistics, methods and computing
- social work
- socio legal studies
- sociology.

6.2 Strategic values

The UK Data Service acquires data to meet three strategic values:

- Potential secondary use and analysis for research: to enable researcher access, asking new questions, undertaking a restudy or follow-up study, or undertaking methodological work, to materials originally created to inform and support research, or to new and novel forms of data not directly collected for academic purposes but of interest to the social scientist.
- Teaching and learning use: to provide access to data that have been created or repurposed specifically for teaching purposes to enhance use of the collection and support skills development.
- Replication and validation of research: to provide researchers with access to original research, which allows checks for consistency, validation of sources and validation and replication of research outputs, both published and unpublished.

6.3 Out of scope materials

Where materials fall outside of scope, efforts are made to determine whether another archive, data centre or responsible institutional repository is more appropriate for the dissemination, curation or preservation of these data.

There are three criteria for not accepting data:

- Legal and ethical issues: there are insurmountable rights management issues e.g. consent, Intellectual Property Rights, copyright and Data Protection issues which are unable to be satisfactorily resolved and where full use of the data would not be possible without infringing legislation.
- Sample size: there are problems with the sample size, the data are very localised or are of peripheral interest to social scientists.
- Lack of sufficient contextual materials to enable re-use.

7. Selection and appraisal criteria

The UK Data Service takes a proactive approach in identifying collections to meeting the needs of its designated user community. The Policy is supported by selection and appraisal criteria that are both transparent and accountable to enable monitoring of processes and planning. Materials will be rigorously reviewed to determine if they may be viably managed, preserved and distributed. The criteria used to review materials consider:

- *Relevance to the remit:* The content of the materials meet the ESRC's current strategies and are within the 'Scope of Collections' (section 4). This includes being high quality, authoritative, reliable, up-to-date with good temporal and spatial coverage, with long and consistent time series (as appropriate), and strong opportunities for comparable research.
- *Scientific or historical value:* The materials are considered important resources either for research purposes and/or meet current demands in research trends and/or there is research and/or teaching activity in the subject area covered by the data.
- *New sources or types of data:* The materials fill a gap in the collection, may supplement areas of the collection, meet new research, such as interdisciplinary, drivers and/or have been requested by users.
- *International value.* The materials are considered to be important for international research.
- *Uniqueness/risk of loss:* The materials form a rare and complete resource of value to the research community and/or are at risk of loss if not acquired; data collections which are costly to reproduce or where continued access to data is threatened; data are not generally available in any other form e.g. paper.
- *Usability/redistribution/operational benefits:* The materials have been well prepared according to data format/documentation and ethical/legal guidelines and fall within Scope; accession to the UK Data Service makes the resources more accessible. Where data is made 'Open' by the data owners, the UK Data Service may continue to make accessible versions of the data because it fits well with the portfolio of other data.
- *Replication Data and Resources:* Materials and related outputs required to reproduce published or unpublished research, or undertake a restudy.

The Collections Development Selection and Appraisal criteria document sits alongside this Policy to enable implementation of the Policy suggestions or materials identified for inclusion in the collection can be recorded, and their ingest managed in a transparent and accountable

manner.¹²

8. Curation categories

The policy of the UK Data Service is to classify data collections in four broad categories:

1. Data collections selected for long-term curation. These data will have long-term secondary analysis potential, and are reviewed, quality checked and enhanced. These collections are likely to be made available for download or order, or accessible via online access tools.
2. Data collections selected for “short-term” management. These collections will not (initially) be retained for long-term preservation, rather they will be backed-up (i.e. bit-level preservation only), are reviewed and quality checked and made accessible and discoverable through online access tools or repository software (ReShare).¹³
3. Data collections selected for ‘delivery’ only, e.g. where data from third parties are accessed via APIs/web services and delivered to end users via a UK Data Service interface. Issues such as level of trust in owner, what documentation/metadata are required, and how rights/registration are handled would need to be agreed.
4. Data collections selected for “discovery” only. These collections will not be brought formally into the UK Data Service collection, they will exist only in other (institutional) repositories, but the UK Data Service will harvest (or in exceptional circumstances, create) metadata records to increase their discoverability.

Data collections may initially fall into categories two or three, but be moved to category one at a point of time in the future.

9. Licence and access framework

To support FAIR¹⁴ data sharing, the implementation of the Policy, the UKRI ESRC Research data policy and the UKRI ESRC Strategic Delivery Plan the UK Data Service implements a robust, three tier licence and access framework including the provision of a Trusted Research Environment for personal data, personal information and/or sensitive data commercially or otherwise.

Data are classified according to their level of detail, sensitivity and confidentiality and appropriate data handling and access safeguards are put in place as follows:

¹² <https://ukdataservice.ac.uk/app/uploads/cd234-collections-appraisal.pdf>

¹³ These include UK Data Service online data access tools, Nesstar, QualiBank and UKDS.Stat.

¹⁴ <https://www.go-fair.org/fair-principles/>

Open data

- Suitable for data that are neither classified as Personal Data nor Personal Information and with no residual risk of disclosure or where consent to share Personal Data or Personal Information as Open Access is in place.
- These are made available to any user without the requirement for registration for download/access.
- Usual licences for re-use are the Open Government Licence¹⁵ and Creative Commons¹⁶ variations.

Safeguarded data

- Suitable for data that are neither classified as Personal Data nor Personal Information but where there is a potential residual disclosure risk. These data are effectively anonymised¹⁷.
- Use of data is subject to the UK Data Service End User Licence Agreement¹⁸.
- These data are made available to registered, authenticated users, and where appropriate, additional conditions and agreements are put in place.

Controlled data

- Suitable for data classified as Personal Information or Personal Data and data that are particularly sensitive, commercially or otherwise.
- Use of data is subject to the End User Licence Agreement and facilitated via the Five Safes framework¹⁹.
- These data are made available to registered, authenticated, trained and accredited researchers, with projects approved by the data owner(s) through a secure connection method.

Deposit licences are negotiated during the acquisition process with data owners. The UK Data Service supports and promotes widespread use of data collections with a key focus on legal and ethical considerations for sharing and using data. There are no costs associated with the deposit or accessing of data for non-commercial and teaching usages.

¹⁵ <https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

¹⁶ <https://creativecommons.org/licenses/>

¹⁷ <https://ico.org.uk/media/about-the-ico/documents/4018606/chapter-2-anonymisation-draft.pdf>

¹⁸ <https://ukdataservice.ac.uk/app/uploads/cd137-enduserlicence.pdf>

¹⁹ <https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/>

www.ukdataservice.ac.uk

collections@ukdataservice.ac.uk

We are supported by the Universities of Essex, Manchester, Southampton, Edinburgh, University College London and Jisc. We are funded by UKRI through the Economic and Social Research Council.